DEVELOPMENT OF A STRUCTURED APPROACH TO MEASURING AUDIO QUALITY OF MOBILE RADIOS.

A thesis submitted in partial fulfilment of the requirements for the Degree

of Master of Engineering

at the University of Canterbury

by James D Collett

University of Canterbury

2009

Table of Contents

List of Figures

Acknowledgments

I would like to acknowledge my supervisors John Pearse, Alan Murray, Emily Lin and Philip Bones, whose support and guidance throughout the project has been much appreciated. Also, thank you to Andrew Ashton and Tony Berggren, whose support at the beginning of the project and with technical knowledge and the administrative side was very helpful.

I would like to thank Tait Electronics Limited and The Acoustics Group at the University of Canterbury for providing a pleasant work environment with helpful colleagues and access to resources. Also thank you to the Foundation For Research, Science and Technology for their support throughout the duration of the project.

I would finally like to thank my wife Becky Collett for all her help, support and encouragement over the past few months.

Abstract

In a communication system, audio quality is one of the parameters by which the end user defines the value of a product. This thesis examines the term audio quality, breaking it down into two subsidiary components, speech quality and speech intelligibility.

One key goal in assessing audio quality is quantifying it in an accurate and repeatable way. As a part of this project a system was developed that achieved this goal. The system was then used to evaluate a number of existing products based on speech quality and intelligibility. Using these results the relationship between the two parameters was investigated. Investigations were also conducted in order to determine and quantify the effect communication systems have on perceptual speech parameters, and examine the relationship between them and speech quality and intelligibility.

Using the testing systems developed a possible method of audio quality optimization was investigated and tested. The analysis methods that were incorporated into the test suite included the Perceptual Evaluation of Speech Quality, the Speech Transmission Index, vowel space analysis and segmental, psychoacoustic based methods. The testing incorporated a number of different handheld portable radios as speakers.

The results obtained from the testing showed that the test suite that was developed operates well, providing meaningful and repeatable results. The investigation into phonetic analysis had varying levels of success, laying the foundations for future improvements. The results culminated in the successful implementation of an audio quality optimization equalizer, which was used to determine the relationship between a number of filter parameters and speech quality.

1 Introduction

In all speech based communications systems, audio quality is becoming an increasingly important parameter, which users employ as a basis for their opinion regarding the quality of a product. Audio quality can be influenced by many variables, both on a perceptual and an analytical level. However, developing tools to measure audio quality can greatly improve the development process as products can be compared using more than the user's subjective opinion.

Tait Electronics Ltd instigated the project in conjunction with the University of Canterbury in order to develop a test suite for objectively measuring audio quality, and investigating parameters that affect quality. The technologies and conclusions drawn from the work are applicable to both the improvement of existing products and the development of future ones. The project was supported by Technology New Zealand via their Technology Fellowship program, encouraging innovative companies to involve postgraduate students in their research and development programmes.

The objective of the project was to design and implement a system that could measure audio quality, such that the results correlated with those of subjective testing. The system was to be incorporated into a test suite designed to test different radios and quantify their audio quality. Using the test suite, a number of radios were to be compared in order to determine the effect of software and hardware changes on the audio quality. The test suite was also to be used to investigate methods of improving the audio quality.

The methods used for measuring audio quality were high level objective measures, the Perceptual Evaluation of Speech Quality [1] and Speech Transmission Index [2], as well as other lower level measures. These measures included perceptual based methods such as vowel space and segmental analysis, which consider the composition of the speech signal, mainly in the frequency domain, and see how the radio's transmission process changes the signal.

A number of radios and loudspeakers were tested using the analysis methods described above. The radios were from the same series, each one representing a different stage in the development process. The results from the testing were then validated using variations of the test suite. This was used to determine the effectiveness of the audio quality measures.

Section 6 addresses the feasibility of implementing an equalizer in the radio that can be iteratively tuned to optimize audio quality. This section describes the implementation and some initial results of such a filter when used with one of the radios.

Two terms which are used extensively in this paper are audio quality and hand held radio. These terms are defined below.

1.1 Audio Quality

One of the most important definitions in the project is the definition of the term audio quality. For this project the term was split into two components, speech intelligibility and speech quality. The definition of these two terms is,

Speech Intelligibility – How easy a segment of speech is to understand.

Speech Quality - How pleasant a segment of speech is to listen to.

Both of these terms can be measured subjectively using test subjects and specific test routines. Speech intelligibility can be measured using the Diagnostic Rhyme Test (DRT) defined in ANSI standard S3.2 – 2009 [3]. Listeners are played a series of test signals and asked to write down what they hear. The system is then graded on the percentage of the words correctly heard, and from this an intelligibility measure is determined.

Speech quality is much more subjective than speech intelligibility; it essentially represents how the user feels about the received signal. Speech quality can be biased by factors such as user expectations, the range of qualities used in experiments, preferences of the subjects and context of the experiment [4]. Quality can be measured using subjective testing. One popular method of this is the Mean Opinion Score (MOS), described in the ITU.T P.800 standard [5, 6]. In this method, subjects are played a number of audio files and asked to rate the quality of the system on a scale of 1 to 5, 1 being bad and 5 being excellent. A public telephone network typically has a MOS of 4.3 and a change of 1 MOS is easily noticeable, ½ MOS is audible and ¼ MOS is just noticeable [7].

The subjective nature of these two measures makes them difficult to measure objectively, but both are also extremely important in products that involve audio reproduction. Hand held radios fit into this group, where users demand high intelligibility in the noisy environments, as well as low noise speech signals in quiet environments.

1.2 Handheld Radios

In this project a hand held radio is a device used for the receiving and sending of Radio Frequency (RF) signals from or to other radios, with the RF signal normally containing a speech signal. All of the testing completed in this project is looking at the radio used as a receiver.

The receiving portion of the radio consists of a number of key component blocks based on the path the audio signal takes, as shown in Figure 1. The first of these is the rf receiver, responsible for receiving the frequency modulated carrier signal from the transmitting radio. This signal is then demodulated and converted down to base band. The output of this block is an audio signal which is sent to the Digital Signal Processor (DSP).

The DSP is responsible for a number of processes that involve the manipulation of the audio signal. These processes include filtering and amplitude adjustment (depending on the strength and quality of the received signal), as well as a number of other processes including volume control. The output of the DSP is an analogue audio signal which is connected to the Power Amplifier (PA).

This amplifier converts the low power audio signal from the DSP to a high power signal which is then routed to the speaker. The speaker is the final stage in the audio production process in the receiving radio, where the signal is converted from electrical to acoustic energy.

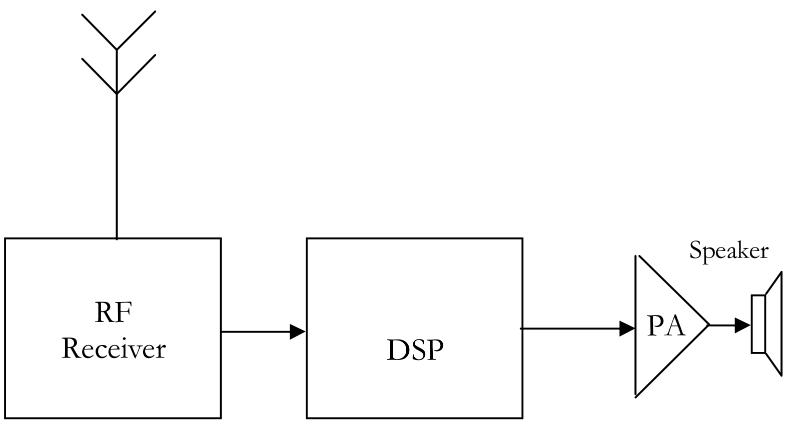


Figure 1: System diagram of a generic hand held radio receiver.

2 Analysis Methods

In order to objectively quantify speech quality and speech intelligibility a number of different analysis methods were used. The methods can be classified into two groups, the high level psychoacoustically based methods and the lower level, phonetic or segmental based analysis. The high level measures used were the Perceptual Evaluation of Speech Quality (PESQ) [1] and Speech Transmission Index (STI) [2]. The lower level methods consisted of measures of vowel space and singer power ratio for vowels and speech moment analysis for consonants.

2.1 Perceptual Evaluation of Speech Quality

This section describes the background and implementation of the PESQ method, compares the PESQ method to other analysis methods and discusses the advantages of using PESQ over these other methods.

2.1.1 Background

There are many objective test methods available for assessing the audio quality of a transmission system. The purpose of most of these test methods is to allow for a systematic and efficient way of objectively approximating the audio performance of a system based on a subjective scale, such as the Mean Opinion Score (MOS) [5, 6]. There are two different types of objective test methods, intrusive and non-intrusive [7]. Intrusive testing uses both the input and output signals for analysis, while non-intrusive testing uses just the output signal. Intrusive testing is simpler to implement and gives results that better correlate with subjective scores.

Since both input and output signals were available an intrusive method was implemented for this project.

The basic concepts of intrusive test methods are similar and can be split into two processes, preprocessing and a distance measurement [7]. The preprocessing procedure involves extracting parameters, such as phase delay and amplitude variations, then transforming the signals into a relevant domain. The domains used are normally temporal, spectral or perceptual. Following the preprocessing and transformation a distance measurement procedure is performed to assess how much the output signal differs from the original input signal for each of the parameters under consideration. A combination of all these parameters is then mapped using a mapping function to give a result which correlates with the MOS score.

The PESQ algorithm is a perceptual domain measure, based on the psychophysics of human hearing. Using this technique, analysis is more credible and likely to produce results that better reflect those of subjective testing. The algorithm is therefore suitable for using with digital mobile radios, as the results are independent of the coding algorithm used. If spectral analysis is used, the result may not detect errors caused by the coding method, as the same error can be a result of both the analysis and coding method.

The PESQ measure, described in ITU 862 [1, 4] was chosen to replace the older Perceptual Speech Quality Measure (PSQM), described in ITU 861[8] in 2001. The newer PESQ measure has been shown to have a significantly higher correlation with results from subjective tests. The major difference between the two test algorithms is the inclusion of an asymmetric and symmetric distortion term in PESQ. These parameters reflect the different types of distortion present in the degraded signal, with each one having a different effect on the speech quality.

2.1.2 Description of Method

Being an intrusive analysis method, PESQ works by comparing the original test signal to one that may have been degraded by the system under test. This is primarily done by performing a perceptual transform on both the reference (input) and degraded (output) signals and determining an asymmetric and symmetric disturbance measurement. Due to the perceptual nature of the algorithm, the parameters being compared represent those that a test subject would use to subjectively determine the audio quality. The analysis is done on a frame by frame basis with all measurements eventually combined to give a single number. The disturbance measure involves a weighted summation, the coefficients of which have been determined through extensive subjective testing to achieve the highest correlation between the PESQ and MOS test results.

There are a number of processes performed on the signals prior to the auditory transform. These include level alignment, input filtering and time alignment, as shown in Figure 2. The PESQ algorithm assumes a constant volume level throughout the playback and recording of the signal. A constant phase delay is applied to both the reference and degraded signals so that they are both aligned with one another.

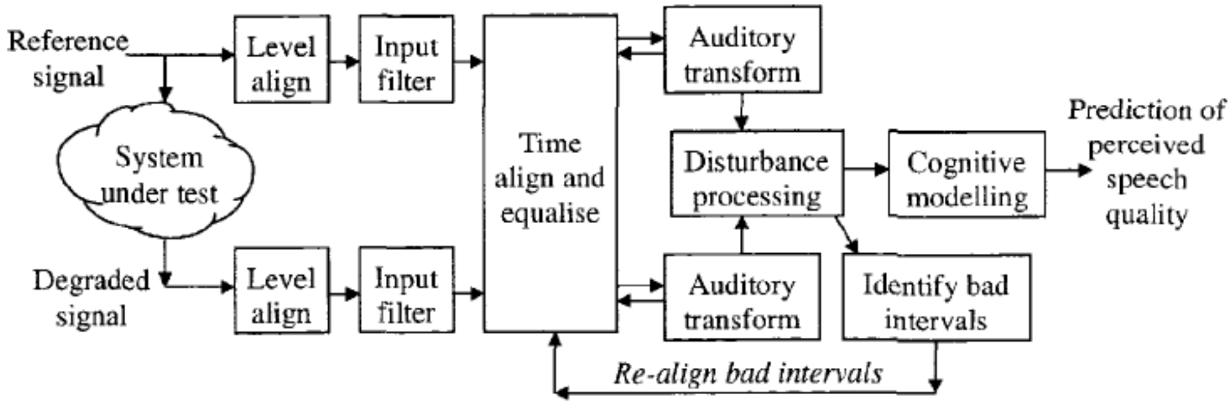


Figure 2: Overview of the PESQ analysis process, taken from [7].

The PESQ algorithm assumes that the subjective listening level of the signal is 79dB SPL at the point the signal was recorded. The level alignment process applies a gain to the reference and degraded signal in order to achieve the same standard level. This means that the PESQ result is theoretically independent of sound pressure or volume level of the test signal. Therefore PESQ cannot be used for measuring the change in quality due to changes in volume.

The time alignment process involves a number of steps, including filtering the signals to emphasise perceptually important parts, completing a series of signal and utterance based delay estimations, and deriving a fine delay identification from the correlation histogram of each utterance. Finally, utterance splitting and realignment is performed in order to determine the delay changes during transmission. The result of this process gives a frame by frame delay estimation of the speech signal which is then used in the auditory transform.

The auditory transform is a psychoacoustically based transform which gives a representation of the signal based on the perceived loudness with respect to time and frequency. The transform is split into a number of stages, resulting in a perceptual loudness value for 42 frequency bands with respect to time. The transform involves usage of the Bark [9] and Sone [10]frequency spectrum and a series of frequency based gain equalisations. The Bark and Sone are loudness measures which take into account the psychoacoustic properties of the human auditory system.

The difference between the reference and degraded signal yields a final PESQ score. This difference is determined via a number of different parameters, including the asymmetric and symmetric disturbances which are determined through a process of segment deletion and masking. The PESQ score is then calculated using Equation (1)

$$PESQ_{MOS} = 4.5 - 0.1\, d_{SYM} - 0.0309 d_{ASYM} \tag{1}$$

where d_{SYM} and d_{ASYM} represent the symmetric and asymmetric disturbance components. The symmetric disturbance component is a quantity defining the distance between the original and degraded signals are, with the exclusion of an asymmetric term. The asymmetric disturbance component is calculated from the ratio of the stabilized bark spectral densities of the original and degraded signals [11].

The PESQ algorithm has been calibrated using large databases of test signals and subjective test results. The weighting coefficients within the algorithm have been adjusted in order to result in the highest correlation between the PESQ results and those of subjective tests.

2.2 Speech Transmission Index

This section describes the background and implementation of the Speech Transmission Index (STI) algorithm, based on IEC 60268[2]. This method was used in parallel with the PESQ method to objectively examine the audio quality of various products during this project.

2.2.1 Background

The STI [2] was originally developed in the late 1960s to provide an objective method for determining the transmission quality of radio transmission systems. The STI provided a faster and cheaper alternative to the existing subjective testing [12]. The Speech Transmission Index now exists in many different forms, the use of which is determined by the application. These variations include Room Acoustical Speech Transmission Index (RASTI), Speech Transmission Index for Telecommunication Systems (STITEL) and Speech Transmission Index for Public Address Systems (STIPA).

The test is performed by transmitting a test signal of modulated, filtered noise (described in full below) through a transmission system and analyzing the output. From this analysis a number between 0 and 1 is calculated to reflect the intelligibility of the transmission system, where 0 represents minimal intelligibility and 1 high intelligibility.

In this study, the Speech Transmission Index was obtained based on rules specified in IEC 60268-16 [2].

2.2.2 Test Signal

The test signal used for deriving the STI is created through modification of a 10 second white noise sample. The signal is filtered into 7 distinct octave bands in the range of 125 to 8000 Hz and the amplitude of each octave band is adjusted by a weighting factor, shown in Table 1. These weightings approximate the same octave band spectral content as the long term spectrum of human speech. In this study, only weightings representing the male speech were used in order to reduce the number of measurements.

After the weighting process, each weighted octave band signal is amplitude modulated by the modulating function in Equation (2), where f_m represents one of 14 modulating frequencies. The modulating frequencies used are 0.63, 0.80, 1.0, 1.25, 1.60, 2.0, 2.5, 3.15, 4, 5, 6.3, 8, 10 and 12.5 Hz.

Table 1: Octave band weightings for the STI test signal

Octave Band(Hz)	125	250	500	1000	2000	4000	8000
Males :	2.9	2.9	0.8	6.8	12.8	18.8	24.8
Females :		5.3	1.9	9.1	15.8	16.7	18.0

$$\text{ModFunc} = \sqrt{1 + Cos2\pi . f_m t} \tag{2}$$

The result of this process is 96 different test signals which have sinusoidal varying intensities, as shown in Figure 3.

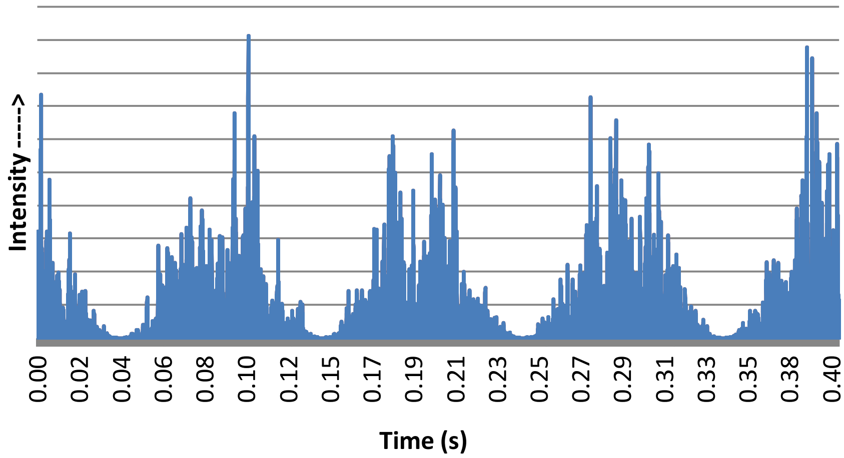


Figure 3: Intensity of the STI test signal in the 2000 Hz octave band and modulated at 10 Hz as a function of time.

2.2.3 Description of Method

The calculation of the Speech Transmission Index can be broken down into 5 individual calculations to yield measures of Modulation Index, Effective Signal-to-Noise Ratio, Transmission Index, Modulation Transfer Function and Speech Transmission Index. The Modulation Index is calculated from the recorded output signal and then used to calculate the Effective Signal-to-Noise Ratio (Effective SNR) and the Transmission Index. The Transmission Indices for individual modulating frequencies are summed to derive Modulation Transfer Indices for each octave band. The Speech Transmission Index is a weighted summation of all the Modulation Transfer indices. Each of the 5 indices is explained in more detail in the following.

Modulation Index

The Modulation Index generally refers to the ratio of the intensity of the original modulating signal to that of the noise that has been added via transmission. According to Steeneken and Houtgast [12], the Modulation Index is defined by Equation (3), where I_{test} is the average intensity of the modulating signal and I_{noise} is the intensity of the noise added as a result of the transmission.

$$m = \frac{I_{test}}{I_{test} + I_{noise}} \tag{3}$$

where:

$$I_{noise} = I_{min} \qquad and \qquad I_{test} = \frac{I_{max} - I_{min}}{2}$$

and I_{max} and I_{min} are the maximum and minimum intensity values, respectively, of the detected envelope of the recorded test signal. Consequently, the Modulation Index can be rewritten as shown in Equation (4).

$$m = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{4}$$

In order to the find values for I_{max} and I_{min}, envelope detection needs to be performed on the recorded test signal. Due to the random nature of the carrier signal and the information known regarding the modulating signal, a method was devised to retrieve the modulating signal from the received waveform, as shown in Figure 4.

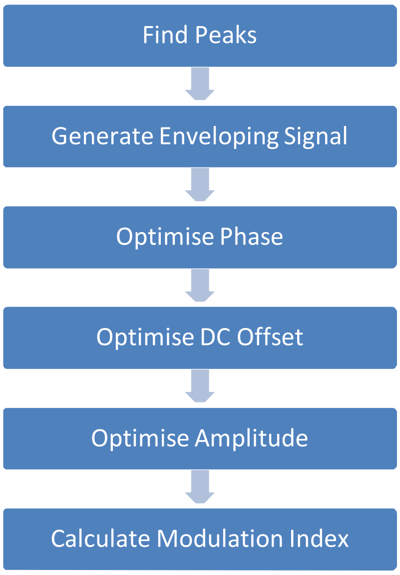


Figure 4: Process for calculating the modulation index.

This method involves generating a sine wave with a frequency matching that of the modulating signal. A cross-correlation procedure is then performed on the intensity peaks derived from the recorded test signal to correct the generated sine wave for phase, DC offset, and amplitude. This correction process is achieved by iterating values of the parameter in question and calculating the cross-correlation of the generated sine waves with the incremented parameter with respect to the intensity peaks of the recorded test signal. The parameter that gives the largest cross-correlation result is used to implement the final enveloping function. The result of this process is a pure sine wave which matches the amplitude, phase, and DC offset of the received envelope, as shown in Figure 5. A similar cross-correlation process is described by Steeneken and T. Houtgast [13] in one of the earlier descriptions of the STI.

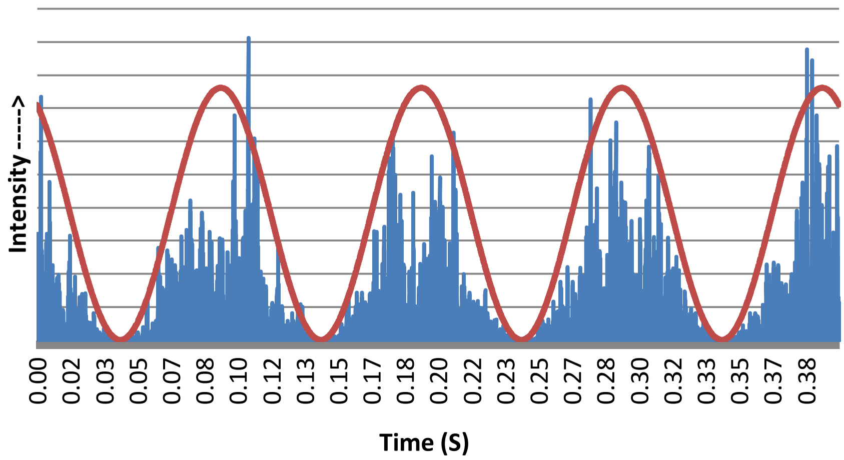


Figure 5: Intensity of the STI test signal and the calculated envelope as a function of time.

Once the enveloping function has been determined, values for I_{max} and I_{min} can be determined to calculate the Modulation Index.

Effective Signal-to-Noise Ratio

Once the Modulation Index is obtained, the Effective Signal to Noise Ratio (SNR) of the test signal can be computed. The Effective SNR is defined as the ratio of the test signal intensity to that of the added noise (see Equation (5)). According to ISO 60268-16 [2], the Effective SNR is defined by Equation (6), where $m_{k,f}$ is the Modulation Index of band k and modulating frequency f.

$$SNR_{k,f} = 10log\left(\frac{I_{test}}{I_{noise}}\right) \tag{5}$$

$$SNR_{k,f} = 10Log\left(\frac{m_{k,f}}{1 - m_{k,f}}\right) \tag{6}$$

The Transmission Index

The Transmission Index is a normalised version of the Effective SNR. It is defined by Equation (7), where $shift$ is 15 dB and $range$ is 30 dB according to [2].

$$TI_{k,f} = \frac{SNR_{k,f} + shift}{range} \tag{7}$$

The Modulation Transfer Index

The Modulation Transfer Index is the average Transmission Index of each octave band. By averaging all the Transmission Index values (Equation (8)), seven Modulation Transfer Indices can be obtained.

$$MTI_k = \frac{1}{14}\sum_{f=1}^{14} TI_{k,f} \tag{8}$$

The Speech Transmission Index

The STI is calculated by using a weighted summation of the Modulation Transmission Index values from each band. Equation (9) defines the weighted summation, where $MTI_n$is the Modulation Transmission Index of band n, and α_n and β_n are the weighting factors for octave band n, as described inTable 2.

$$STI = \sum_{n=1}^{7} \propto_n MTI_n - \sum_{n=1}^{6} \beta_n \sqrt{MTI_n \times MTI_{(n+1)}} \quad \textbf{(9)}$$

Table 2: Octave band summation weighting factors for female and male testing. [2]

Octave Band (Hz)		125	250	500	1000	2000	4000	8000
Males	α	0.085	0.127	0.23	0.233	0.309	0.224	0.173
	β	0.085	0.078	0.065	0.011	0.047	0.095	-
Females	α	-	0.117	0.223	0.216	0.328	0.25	0.194
	β	-	0.099	0.066	0.062	0.025	0.076	-

2.3 Vowel Space Analysis

2.3.1 Background

The vowel space is a two-dimensional area bounded by the first two formant frequencies of the extreme vowels, when plotted with Formant 1 and Formant 2 frequencies on the x and y-axis. Research has been conducted investigating the relationship between vowel space and speech intelligibility for normal talkers as well as people suffering from motor function disorders. By using synthesized vowel sounds with controlled formant frequencies, a test was designed in order to determine the effect the radio's transmission system has on the vowel's formants.

Vowels are created by exciting the vocal tract with the glottal tone, a quasi-periodic signal generated by the larynx. Each vowel has a unique vocal tract configuration, which creates resonating cavities of different sizes. The excitation of these cavities causes resonant peaks to form in the spectral content of the glottal tone and these peaks are the formants of the vowels.

To be able to compute the vowel space, the formant frequencies need to be computed. The Fourier Transform reveals the frequency content of the vowel sound in Figure 6, but, due to the nature of the glottal tone, the harmonics present may not fully excite the resonant cavity. This means that the formant frequency can sometimes be hard to find using the Fourier Transform, especially for speakers with high fundamental frequencies.

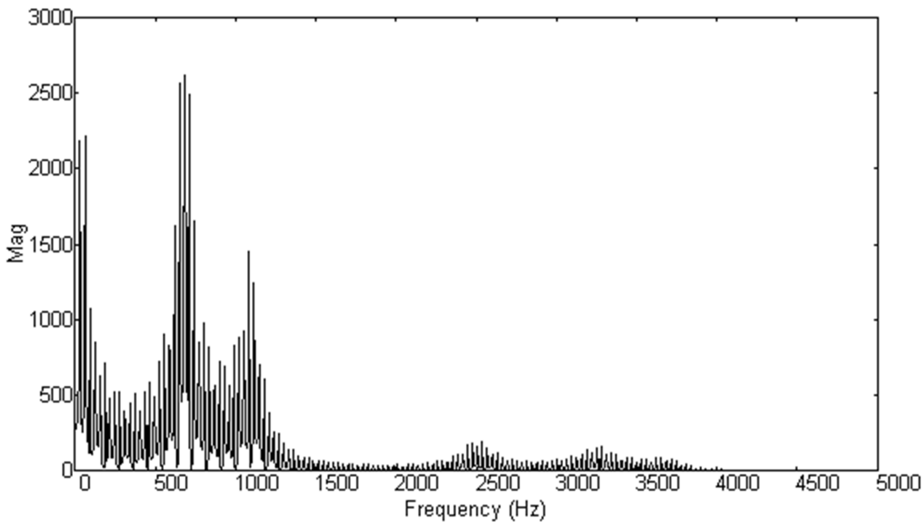


Figure 6: Fourier Transform of the /a/ vowel, with a low frequency.

A popular method used for finding formant frequencies is Linear Predictive Coding (LPC) [14-16]. This method uses prior knowledge about the creation of the vowel sound along with recursive algorithms to find the frequency response of the resonating chamber, in this case the oral cavity. The output signal of a speaker pronouncing a vowel sound can be defined as the output of a linear causal filter $h(k)$, representing the resonant properties of the vocal tract, excited by a white noise signal $w(n-k)$, representing the glottal tone.

The output signal x(n) is defined as:

$$x(n) = \sum_{k=0}^{\infty} h(k)w(n-k) \tag{10}$$

By defining the linear causal filter to be an all pass filter with coefficients $A(z)$ and order p,

$$H(z) = \frac{1}{A(z)}$$

equation 10 can be defined as an auto regressive function,

$$w(n) = x(n) + \sum_{k=1}^{p} a(k)\, x(n-k)$$

The Levinson-Durban recursion procedure [17] is one of the methods available to derive the filter coefficients. The Levinson-Durban algorithm was implemented in `Python` in order to obtain the LPC functionality in a configurable environment, shown in section 10, Appendix B. Using these coefficients, the frequency response of the vocal tract can then be calculated and plotted, as shown in Figure 7. The major difference between the result of the LPC analysis and the Fourier Transform is the smoothing of the spectral content. The LPC method is one of the more popular methods used for formant-based analysis of speech, having been used by a number of authors [14-16, 18].

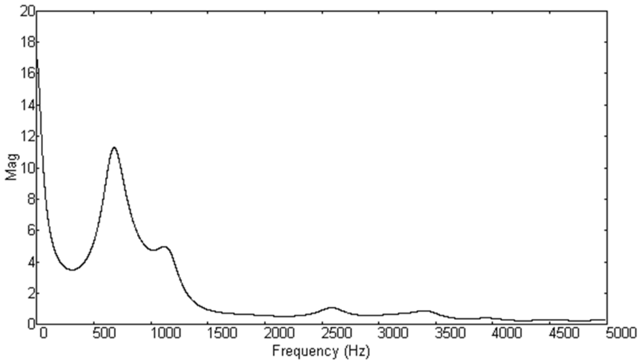


Figure 7: LPC transform of the /a/ vowel.

The first two formant frequencies of all the vowel sounds can be plotted in a two dimensional space, as shown in Figure 8. The area covered by this two dimensional space has been shown to be related to speech intelligibility [15]. There have been a number of ways used to quantify the area covered by the vowels, mainly differentiated by the vowels being used to define the apexes of the vowel space. The main combinations that are used are the /i/, /a/ and /u/ vowels [15, 16], which creates a triangle, or /i/, /a/, /æ/ and /u/ [14, 18], which creates a four sided polygon. The vowels used to define the area's shape are called the corner vowels.

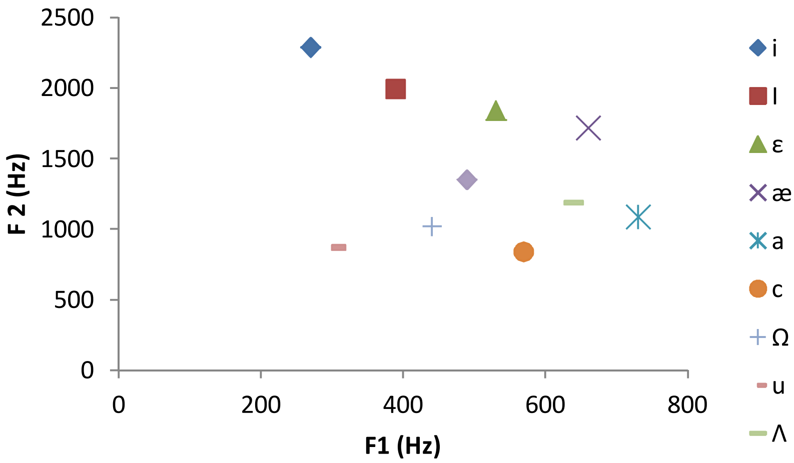


Figure 8: Distribution of the vowel sounds in the F1 - F2 space, formant frequencies are for the average adult male speakers [19].

There have been a number of studies done with the intention of finding a relationship between vowel space and speech intelligibility. The idea behind the proposed relationship is that the larger the vowel space, the more variation between the individual vowel sounds and thus, the easier it is to differentiate between different vowels. Huei-Mei, et al[15] performed listening tests to examine the effects of reduced vowel space on the listener's perception of the vowels. A number of synthesized vowel sounds, where the formant frequencies were controlled, showed that the test subjects found the test cases with the larger vowel spacing to be better examples of vowel sounds.

Testing was also done using real human speech looking for the same relationship. A positive correlation was found between intelligibility and vowel space for speakers who suffered from Amyotrophic Lateral Sclerosis (ALS) [15]. It was determined that this relationship was a result of the test patients lacking the motor skills required to shape their vocal tracts into the maximum and minimum positions. This reduction in movement results in the formant frequencies of the vowels being closer together and thus, a smaller vowel space.

Another study investigated the link between speaking rate, intelligibility and vowel space for both neurologically intact people and ALS sufferers [18]. The subjects were asked to speak a different way for each test (normal, fast and clearly). The study found a positive correlation between speaking rate and vowel space for both neurologically intact and Amyotrophic Lateral Sclerosis speakers.

There has also been work done looking at the acoustic cues, namely vowel space, of different talkers in an attempt to quantify what makes some people more intelligible then others. In these studies subjects were not known to suffer from any motor skill disorder and were asked to speak normally. One study performed subjective tests on the talkers to determine each talker's intelligibility, then performed vowel space analysis on each talker's speech [14]. No correlation was found between the Euclidian area covered by the three corner vowels and the talker's intelligibility. However, when the dispersion of the formant frequencies from a central point in the two dimensional F1 F2 space was considered, a correlation was found between the dispersion and intelligibility. This suggests that the area covered by the three corner frequencies is not as important as the dispersion of the vowel sounds in the two dimensional space when normal talkers are being considered.

The vowel space of talkers that naturally talk at different rates was also examined, to see if fast talkers substitute vowel space for speed [14]. This work revealed no correlation between the vowel space of the talkers and their natural speaking rate and the two groups also had similar absolute formant frequencies. This suggests that the distribution of the vowels in the F1 F2 space is approximately the same within groups of talkers, and therefore must be an important aspect of the acoustic cues in speech.

2.3.2 Description of Method

This testing is designed to determine if a person's speech signal would suffer any change in vowel space due to the processes present in the radio's transmission and receiving systems. In order to conduct this test, signals containing vowel sounds were transmitted through the system, with the output being recorded. Both the original and degraded signals were analyzed in order to determine their formant frequencies and thus vowel space. Linear Predictive Coding (LPC) was used along with a formant tracking algorithm in order to determine the formant frequencies, which were then used to calculate the vowel space.

The test signals used were synthesized vowel sounds generated by filtering glottal tone-like signals to obtain a spectral distribution similar to that of a real vowel sound. Each of the vowels being tested had formant frequencies similar to those of the average formant frequencies of the specific vowel sound when spoken in normal conversational speech. The filter with the resonant peaks similar to those of real vowels was excited with a number of different glottal tone-like signals, each with a different fundamental frequency. The result of this was 59 differently pitched test signals for each vowel sound. The advantage of using synthesized vowel sounds instead of real ones was the flexibility and control which comes with generating synthetic test signals; thus the formant frequencies and pitch of the signal were able to be controlled as well. In addition the length of the signal could be easily designed. In real speech, formant frequencies are affected by the preceding and following sounds. Therefore using test signals derived from real speech could result in some variability to the experiment which could cause inaccuracy in the results.

The formant tracking system used is described by McCandless [20] and used in applications requiring real time formant tracking such as public address system enhancement [21]. This method uses peaks from the LPC spectrum, along with a series of rules, in order to calculate the formant frequencies. The vowel signal is divided into frames of the same length, in this case 20 ms, and the formant frequencies of each frame calculated. The algorithm also utilizes the stable nature of vowel formant frequencies. Formants do not tend to jump instantly from one value to another, instead they transition over intermediate values. This means the formant values from the previous frame are used as estimates in order to weight the formant selection within the current frame.

The processing carried out on each frame is identical and can be broken down into a number of steps. In order, these steps are peak finding, filling the slots, removing duplicates, dealing with unassigned peaks, dealing with unfilled slots and finally recording the results, represented by Figure 9. If there are unfilled slots at the end of the process, peak finding is completed again with slightly different parameters in an attempt to extract hidden peaks, and the process is repeated. As previously mentioned, the results from the current frame are used as estimates for the calculation of the formants in the next frame. The process is repeated for each frame in a recorded signal, producing the selected formant frequencies for each frame.

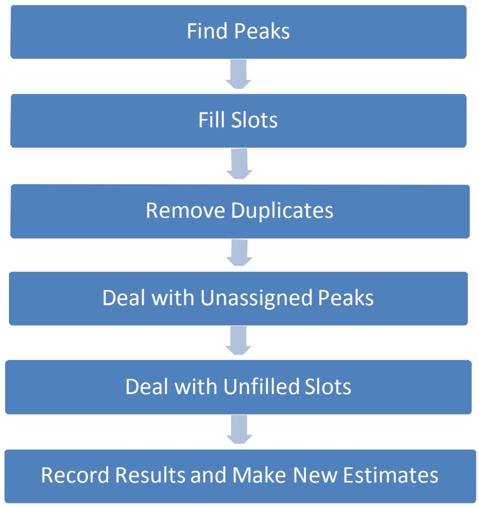


Figure 9: Formant extraction process.

I. Find Peaks

The first part of the algorithm determines the spectral peaks within the current frame. The initial step of this process is to calculate the LPC coefficients; this is done using the process described in the section 2.3.1. Any number of LPC coefficients (p) can in principle be calculated, but for the testing reported here 14 coefficients were chosen. This proved to be sufficient and was also the suggested number in the published description of the algorithm [20]. The magnitude of the transfer function, (z) , of the filter corresponding to the LPC coefficients, A_k, can be calculated by evaluating the magnitude of $H(z)$ around a circle of radius R.

$$H(z) = \frac{A_0}{1 - \sum_{k=1}^{p} A_k z^{-k}} \tag{11}$$

Where:

$$z = Re^{j\left({}^{2\pi n}/_{N}\right)} \tag{12}$$

$$n = 0,1,2 \ldots\ldots, N - 1$$

In equations (11) and (12), N represents the number of frequency divisions within the 0 to f_s range, where f_s is the sampling rate in Hz and R is always chosen to be less than, or equal to 1. For the purpose of this testing the value for N used is 256 which, coupled with a sampling rate of 10 kHz, gives a frequency resolution of 40 Hz. This is also the value stated in the description of the algorithm [20]. The parameter R is initially set to 1 but, as previously mentioned, if there are any empty slots after execution of the algorithm, the peak finding process is repeated with a value of R 0.004 less than that of the previous iteration. The algorithm is then re-run using the newly calculated peaks. This reduction is continued until either all the slots are filled or R is set to a value less than 0.88, at which point it is determined that no formant can be found for that slot in the current frame.

Once the magnitude of the transfer function $H(z)$ has been determined, the four highest peaks below 3400 Hz and above 150 Hz are determined. This is achieved by differentiating

$H(z)$ and searching for the zero crossings in the first derivative for which the second derivative of H(z) is negative. The frequency and magnitude of these peaks, in order of ascending frequency are denoted $P_1, P_{2,} P_3$ and P_4, where P_1 represents the peak with the lowest frequency.

II. Fill Slots

The algorithm starts with four empty slots $S_1, S_{2,} S_3$ and S_4, which represent formants 1 through 3, or possibly 4 respectively. The user also defines four initial estimates of the four formant frequencies, which are used to calculate the formant frequencies in the first frame. Once the highest four peaks have been found, the next step is to fill each of the slots with the peak that is closest in frequency to the estimate for that slot. The same peak can be put into multiple slots if it is the closest to the estimate but at the end of this process all slots are filled with a peak.

III. Remove Duplicates

The next step is to remove any duplicate peaks from the slots. This is done by firstly, searching for duplicates then removing all but the strongest candidate, the strongest having the smallest difference between peak frequency and slot estimate.

IV. Deal With Unassigned Peaks

The next step is to assign any unassigned peaks. If a peak P_x is unassigned and slot S_x is empty then P_x is simply put into slot S_x. If S_x is not empty then the amplitude of P_x and the peak in slot S_x are compared and if the amplitude of P_x is less than half the amplitude of the peak in

S_x, P_x is discarded. Otherwise if slot S_{x+1} is empty then the peak in S_x is moved to S_{x+1} and P_x moves into S_x. If S_{x+1} is not empty then S_{x-1} is examined and if empty, the peak in S_x is moved to S_{x-1} and P_x fills S_x. If P_x remains unassigned after this process, it is discarded.

V. Deal With Unfilled Slots

After the assignment of unassigned peaks has been carried out, if any slot remains unfilled, excluding S_4, the slots are emptied and the process is started again with recalculation of the magnitude of the transfer function $H(z)$ using a smaller value for R. When R gets smaller then 0.88, the process is abandoned and the formants for that frame are assumed not to exist.

If the process is successful a minimum of three peaks are returned, representing formants 1 through 3, and maybe 4. These formants are then used as the initial estimates for the next frame's processing. This biases the process to select formants close in frequency to those in the previous frame, effectively exploiting the expected temporal continuity of the formant frequencies.

When the process has been completed, assuming formants were found in each frame, a list of formants is produced for each frame. An average value is calculated for each formant frequency using the values from all frames.

VI. Vowel Space Calculation

For vowel space analysis only the first two formant frequencies of the three corner vowels are required. The calculation of the formant frequencies is done using Equation (13), which is derived in section 10, Appendix A.

$$Area = \frac{ab}{2} sin(cos^{-1}\left(\frac{a^2 + b^2 - c^2}{2ab}\right)) \qquad \textbf{(13)}$$

where;

$$a = \sqrt{(F_{12} - F_{13})^2 + (F_{22} - F_{23})^2}$$

$$b = \sqrt{(F_{12} - F_{11})^2 + (F_{21} - F_{22})^2}$$

$$c = \sqrt{(F_{11} - F_{13})^2 + (F_{21} - F_{23})^2}$$

and F_{xy} is the xth formant frequency of the yth vowel.

This analysis is carried out in order to determine the effect the radio transmission has on the vowel space of a speech signal. This is done by transmitting three synthetic vowel sounds through the transmission system and recording the output. The above analysis method is then used to calculate the formant frequencies of the recorded sounds, which are then used to calculate the vowel space. The advantage of using an automated system for determining vowel space is that it provides a consistent method for quantifying formants. This analysis method could also be used in the future for enhancing the formant frequencies of the speech signal during transmission thus improving intelligibility in high noise environments.

2.4 Segmental Analysis

A number of other phonetic analysis methods were examined in order to determine some of the effects that the transmission has on a speech signal. There are a number of advantages to this type of lower level, perceptually based testing. For example, using these analysis methods new signal processing algorithms can be developed in order to improve these perceptual measures during transmission.

The phonetic analysis methods include Singer Power Ratio and formant 1 and 2 slope. These measures have been found to correlate with the style of a person's voice, specifically different types of singing [22]. By using these measures to examine the effects of the radio transmission an understanding can be gained with relation to how the radios change speech signals on a perceptual level.

The slope between the first and second formant is determined using the formant results calculated by the formant picking algorithm and equation (14).

$$Slope = \frac{|F_2| - |F_1|}{f(F_2) - f(F_1)} \qquad \textbf{(14)}$$

where f() denotes the frequency of the defined formant and || represents the magnitude.

Singer Power Ratio is quantified as the ratio of the amplitude of the highest peak below 1500 Hz to the amplitude of the highest peak above 1500 Hz. This ratio has been shown to correlate with a person's ability to make their voice carry over background noise, specifically in performance situations such as singing. This measure is calculated using the spectral information obtained from the results of the automatic formant extraction process.

2.5 Frequency Response Analysis

In order to determine the frequency response of a number of the devices being tested, white noise was used in conjunction with the Fourier Transform, utilizing the Fast Fourier Transform algorithm. The white noise signal was played though the test system and recorded at the output of the radio as per the normal test procedures. The signal was then transformed into the frequency domain, via the Fourier Transform, in order to obtain the frequency response.

In order to reduce processing time the signal was also converted from a sampling rate of 48 kHz to 16 kHz, allowing a 4096 point DFT to be performed to provide approximately 4 Hz frequency resolution. For this testing no windowing function was used as it was deemed unnecessary due to the length of the signal available for processing. In addition, because the radios have a bandwidth of 300 Hz to 4000 kHz, a 16 kHz sample rate is sufficient for this analysis as it easily encompasses the pass band spectrum as well as allowing for the first harmonic distortion products.

3 Test Suite

The test suite represents the hardware and software tools that were developed in order to test speech quality and speech intelligibility with respect to volume levels of any Tait Electronics Ltd handheld radio. Variations of this test suite are also used in order to test speakers as well as other non Tait Electronics Ltd products.

3.1 Requirements of the Test Suite

A test suite was developed to evaluate the sound quality of the output of a number of different handheld radios. The designed test suite was required to be portable and fully configurable and adaptable to various test conditions. The test suite was comprised of two contributing parts, the physical test setup and the test suite program. The test suite program provides a user interface for configuring test parameters and running tests. The physical test setup consists of the hardware of the test suite which can be controlled by the test suite program.

In order for the test suite to provide repeatable results and be portable independence between the results obtained and the equipment used was required. In order to achieve this, the equipment used needed to be easily replicable and all the variable parameters within the test setup needed to have their effects investigated and resulting settings defined. The test suite program also needed to be transferable between computers whilst producing comparable results.

Due to the large number of test signals and the number of test iterations to be performed, a level of automation was required. This automation not only makes the testing less labor intensive, but also mitigates the effects of human error. The automated nature also means more test iterations can be run for a particular test, improving the accuracy and validity of the results obtained.

3.2 Physical Test Setup

The physical components of the test suite can be broken up into six discrete components: a computer, USB audio interface with separate microphone, oscilloscope, RF transmitter and the Unit Under Test (UUT), with inter-connections shown in Figure 10. These components have been chosen as they provide the required level of configurability for the test suite. They are also off the shelf, unmodified pieces of equipment, which helps make the testing more repeatable. The test setup also utilises an acoustically isolated environment, the anechoic chamber located in the Electrical & Computer Engineering Department, University of Canterbury.

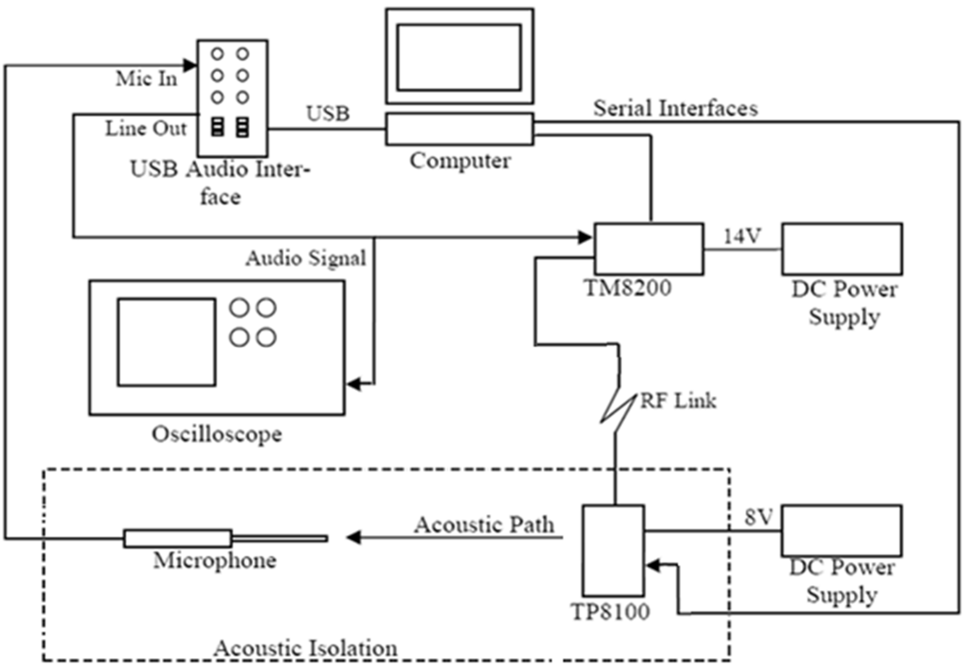


Figure 10: The test setup.

The computer in the test setup runs the test suite program that was developed as part of this project. This program controls the RF transmitter, UUT and USB audio interface as well as all the playback and recording of test files. There is also a small amount of preprocessing of the test signals required, which is also completed by the test suite program. Because the test suite needs to be portable and repeatable, the program is packaged into a single .exe file and requires only a computer with greater than 512 MB of RAM that is running Windows.

The control of the RF transmitter and UUT is done via a Computer Controlled Data Interface (CCDI) [23], which is an interface designed by Tait Electronics Ltd for automated control of their radios via a serial port. The CCDI is used to start and stop transmission of the RF transmitter and change the volume levels of the UUT. There are 25 discrete volume levels on the units being tested and these are all activated using CCDI. The CCDI interface can also be used to emulate any other user input that the radios would usually accept. This flexibility and automatable control increases the potential scope of the testing that the test suite can perform.

The USB audio interface used for the test setup is a Lexicon Lambda Studio and is used to interface all the audio signals to the computer. This device connects to the computer via USB and can stream audio in full duplex. The device has two balanced inputs, to which it can apply 48 Volt phantom power and as two balanced outputs. It also supports 48 kHz and 44.1 kHz sample rates with anything up to 24 bit precision. The datasheet [24] states that the total harmonic distortion of the microphone inputs is less than 0.05% between 20 Hz and 20 kHz as well as having a frequency response between 0.5 and -0.5 dB between 20 Hz and 20 kHz. This helps maintain the integrity of the recorded test signal by minimising the addition of extra noise and distortion. This device is portable between computers, making the test independent of the characteristics of individual computer sound cards.

The RF transmitter for this test setup is TM8200, a mobile radio designed and manufactured by Tait Electronics. An important part of this test setup is that it must approximate real life operating conditions. In order to do this all standard processing must be carried out on the audio signal. By using a mobile radio as the RF transmitter this functionality is obtained and the platform is fully configurable and adaptable because access to the firmware is available. The test suite utilises the TAP IN functionality of the radio's Digital Signal Processor, DSP, to insert the audio signal directly into the audio line. This means that for the purposes of testing, any of the standard processing can be omitted from the audio line in order to determine its effect on the audio quality. For the purpose of this testing the signal is inserted after the Automatic Level Control (ALC) process which, is at the start of the audio line. The ALC process increases the dynamic range of the system by adjusting the gain of the radio's microphone depending on the level of the incoming signal. This process was omitted because it has been shown that the PESQ algorithm overestimates the effects of the ALC on the audio quality.

The oscilloscope is used to monitor the amplitude of the test signal going to the RF transmitter. Because the audio signal is tapped into the audio line after the ALC it is critical that the input level of the test signal is controlled and monitored, as it has a direct effect on the deviation of the RF transmitter. This means if the signal has smaller amplitude than that of a signal being transmitted under normal conditions, the magnitude of the deviations from the carrier frequency will be smaller, and thus the signal to noise ratio of the received signal will also be smaller. On the other hand, if the test signal's amplitude is too large then there is a process on the DSP that limits the signal's amplitude so the output of transmitter does not over deviate. This introduces distortion products to the signal which would not be present under normal conditions, would not be present. In order to utilise the oscilloscope an optimal level for the input signal needed to be determined.

Transmitter Input Level Testing

The optimum signal input level was determined by a number of tests designed to investigate the relationship between SPL, signal level and output deviation. The test setup consisted of a TM8200 mobile radio along with its external microphone, a larger loud speaker capable of producing SPLs above 100 dB, SPL meter, a Hewlett Packard 8920A RF test set and an oscilloscope, as shown in Figure 11.

The speaker, microphone and SPL meter were all situated in an anechoic chamber with the transducer of the SPL meter situated within close proximity of the microphone. The Hewlett Packard 8920A was connected to the RF output of the TM8200 via a 20 dB pad and the oscilloscope was connected to the audio tap out of the TM8200s DSP. The tap location is just after the ALC, as it is in the standard test setup described earlier in this section.

The test procedure involved playing pure tones, with frequency of 1 kHz and 2 kHz, sequentially through the loud speaker at varying amplitudes. Using the SPL meter the SPL presented to the TM8200's microphone was measured for each test case. Concurrently the TM8200 was put into transmit mode, at which point the deviation of the RF output and the amplitude of the audio signal at the specified tap out point were measured.

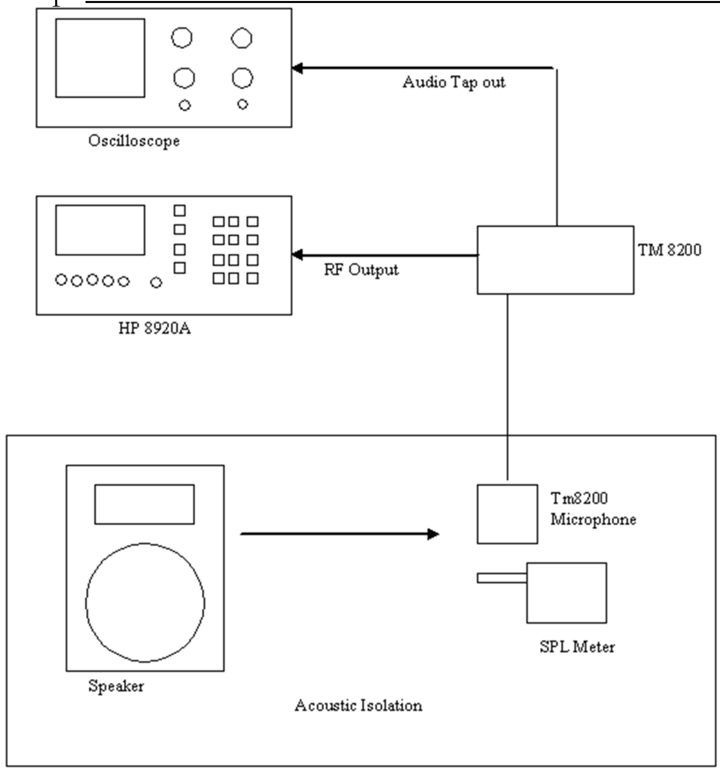


Figure 11: Test setup for investigating the relationship between SPL, output deviation and signal voltage level.

Two test frequencies were used for this testing because there is a process within the audio subsystem called ‘pre-emphasis’ which applies frequency dependant amplification to the input signal. This amplifier has a gain of 0dB at 1 kHz and 6dB at 2 kHz. Using these two testing frequencies the effects of the pre-emphasis can be noted and accounted for in the determination of the optimal input signal level. The difference between the two frequencies, in terms of input level and output deviation, is evident in the results, as shown in Figure 12 and Figure 13.

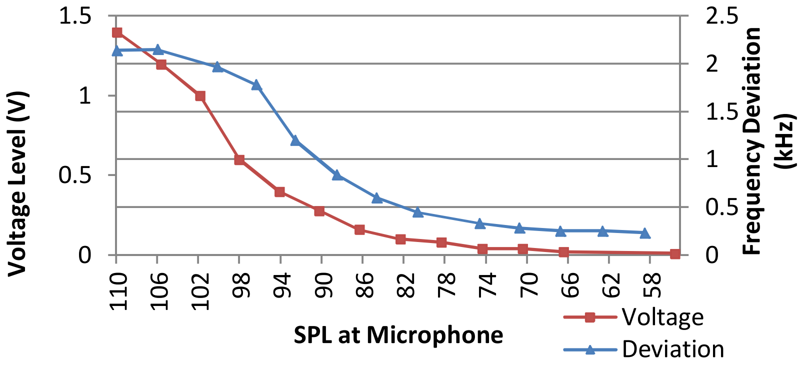


Figure 12: Relationship between SPL of a 1 kHz tone presented to a TM8200's microphone, the voltage level measured at the described tap out point and the frequency deviation of the RF output.

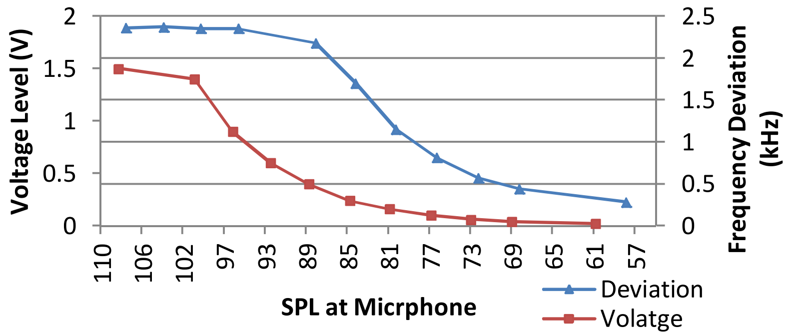


Figure 13: Relationship between SPL of a 2 kHz tone presented to a TM8200's microphone, the voltage level measured at the described tap out point and the frequency deviation of the RF output.

It is worth noting the differences between the relationships found at the two different test frequencies. The output frequency deviations and tap out voltage level generated by the2 kHz tone both increase at a lower SPL then those generated by the 1kHz tone. Both of the parameters being measured also approach an upper limit, after which the rate of increase is reduced. This is due to effects of the pre-emphasis, which was described earlier, and a limiting function which controls the maximum output levels.

To relate these results to the standard test setup, another test was performed relating the input amplitude of the test signal with the test setup in the configuration shown in Figure 10, and the deviation of the RF output was determined at both 1 kHz and 2 kHz. For this test the input signal amplitude was varied but, for each measurement, the output level from the computer was set to either 100% or 60 % of the computer's maximum output capability. This was done so that it could be observed when the limiter started reducing the gain of the signal. This is seen in the results by a reduction in the change in deviation between the 100% and 60 % test, as shown in Figure 14 and Figure 15.

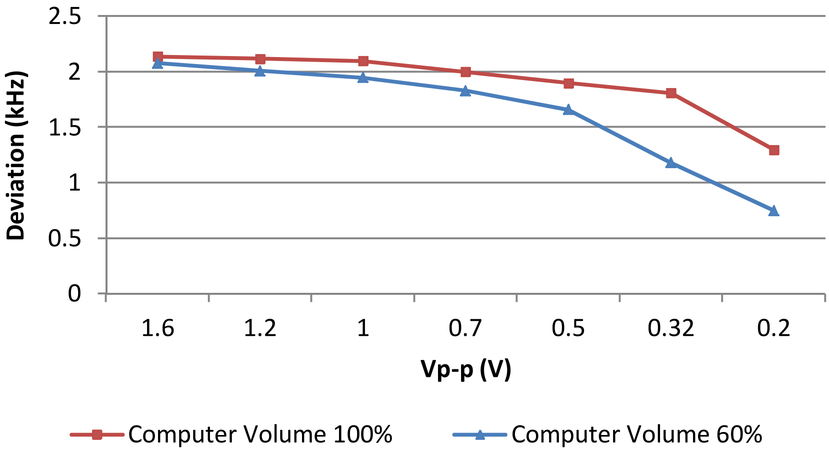


Figure 14: Relationship of frequency deviation measured at the output of the transmitting radio and peak to peak voltage for a 1kHz sine wave.

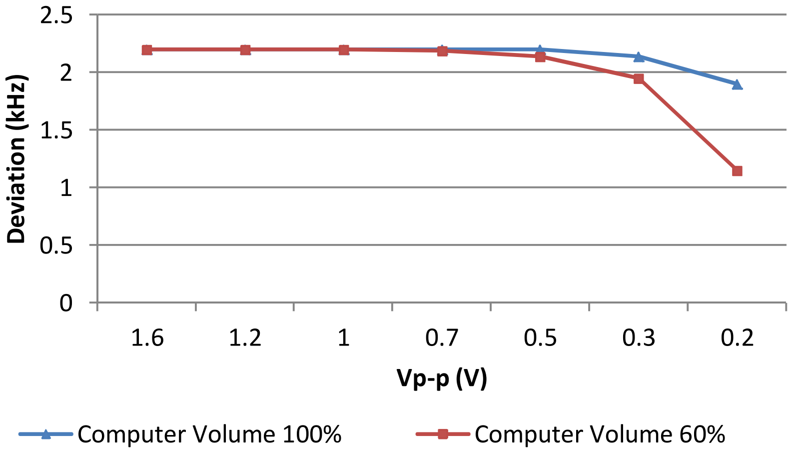


Figure 15: Relationship of frequency deviation measured at the output of the transmitting radio and peak to peak voltage for a 2 kHz sine wave.

From this testing it was decided that the input level of the test signal required a peak to peak voltage of 0.4 volts measured at the tap in point, when the computer is playing a 1 kHz sine wave at a volume level of 100%. Also all audio files were normalized on the computer so that the maximum amplitude within any of the signals represents the maximum volume level that computer can produce. If 0.4 V is traced on Figure 12, Figure 13, Figure 14 and Figure 15 it shows that the test signal, when at its greatest amplitude, will be just activating the limiter for both the 1 and 2 kHz. This not only makes the test represent real transmission conditions, but also makes the test more repeatable by specifying particular voltage levels.

The microphone used in the test setup was a Behringer ECM8000 reference microphone. This microphone is an omni directional, electret condenser microphone with a frequency response of between -2dB and 2 dB between 20 Hz and 20 kHz [25]. The microphone, being a condenser type, required 48 Volt phantom power which is provided by the USB audio interface. This microphone is used for this testing because it provides a relatively flat frequency response within the spectrum of interest, helping with the repeatability of experiments.

The UUT is the name given to the device being tested. For radios for which CCDI is available, the computer controls the volume level via the serial port, as previously described. For radios that do not have this feature, volume levels must be set manually, which reduces the accuracy and repeatability of the experiments but, when combined with some form of output SPL reading, still reveals some relevant and repeatable results.

All the testing for this project was completed in the Anechoic Chamber located in the Electrical & Computer Engineering Department at the University of Canterbury, shown in Figure 16. The room has a floor area of 17.7 m^2, a total suface area of 97.3 m^2 and a room volume of 63.6 m^3 [26]. This chamber was used because it provided a free field acoustic environment with a very low background noise level, (c.f. Appendix C, section 10). Also by performing testing in an acoustically isolated environment, the test is more repeatable and the results more valid as they are not influenced by any significant parasitic noise.

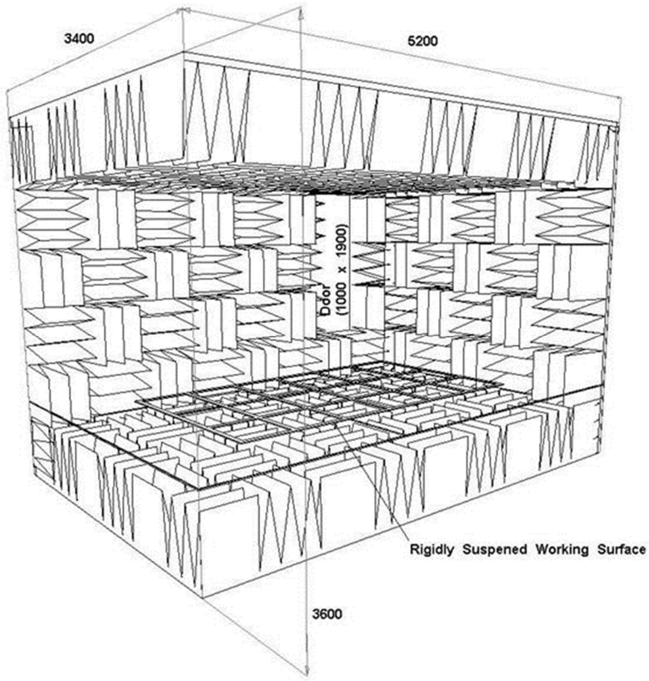


Figure 16: Design and layout of the Anechoic Chamber in the Electrical & Computer Engineering Department at the University of Canterbury.

3.3 Test Suite Program

The test suite uses a graphical user interface from which the program can be controlled. From here the user can select the test files to be played, can select the directories for files to be output to and can initiate preprocessing of the test signals. The user also needs to set up the CCDI options, communication (COM) ports to be used and baud rate. The USB audio interface is also configurable and the options that can be set include sample rate, bit precision and the sound device to be used. Finally, before the test begins the user selects which volumes levels are to be tested. The user then starts the test and the program displays a set of progress bars to indicate test progress.

3.4 Test Process

When the test is running the test suite iterates through a number of processes for each test case. These processes are: start transmission, commence recording, begin playing test file, stop transmission and finally, stop recording. The processes are repeated in a cyclic manner until the test is complete.

3.5 Test Suite Variations

Because the test suite needs to be configurable and adaptable it stands to reason that there are variations to the standard test setup. During this project, other variations of the test setup were used for testing speakers and radios which did not have CCDI. For each of these tests there were slight modifications made to the standard test setup.

3.5.1 Loud Speaker Testing

The objective of speaker testing was to compare a number of different speakers in a number of different configurations using the analysis tools described in section 2. The speakers were to be tested both in a sealed environment, which emulates a configuration similar to that of the radio, and in free field, where the speakers effectively have minimal acoustic isolation between the back and the front of the diaphragm. The speakers were also driven by a radio and a comparatively large, wide band amplifier. The radio was used in order to emulate real world operating conditions and the wide band amplifier removed any degradation due to the radio's sub system.

The normal method for mounting the speakers in the radios involves gluing the driver into the chassis of the radio. This provides a robust, water tight seal ideal for use in the field. However the speaker cannot be easily removed from the chassis and not without causing damage to the speaker. This makes this mounting system impractical for speaker testing. For these reasons another mounting system was devised and implemented. The new system needed to incorporate similar acoustic specifications to those of the radios, namely cavity volume, whilst providing a sealed enclosure. Also to make the new mounting system effective, speakers needed to be able to be easily inserted and removed whilst maintaining their integrity.

The mounting system used for the speaker testing consisted of a molded plastic box and RCA plug. The box has a volume which is similar to that of a radio and had rubber seals around all the joins creating a sealed enclosure. A hole was drilled in one side of the box with a diameter slightly smaller than that of the speaker. The edge of the hole on the inside of the box was machined out to provide seating and alignment for the speaker. The speaker was held in place by a plastic brace across the magnet which was held in place by a bolt on either side of the speaker. The speaker terminals were connected to a female, panel mount RCA plug which was mounted on the side of box. This housing provided a standard, sealed test enclosure that had a similar volume to the radios and could be used for accessing any rear mounted speaker. For the free field testing the back of the enclosure was removed but the speaker remains in the front panel mounting. This meant that the speakers were mounted consistently for the free field testing and thus reducing possible variations in the results.

Figure 17: Photos showing a speaker mounted in one of the boxes used for performing the speaker testing.

In order to test the speakers in the required configurations, changes were made to the physical setup of the test suite. Two variations of the test suite were used to test the speakers when being driven by a radio and comparatively wide band amplifier. The same analysis methods were used and the majority of the testing process remained unchanged.

The radios have the facility to accommodate an external loud speaker or a headset earphone. This functionality was used to drive the speakers for the radio driven tests. A switch was used to swap between the internal speaker and the accessories port so that the post-amplification processes that the audio signals undergo remained the same. This means that the tests reflect the standard operating conditions of the speaker in the radio in terms of degradations to the signal due to transmission, processing and amplification.

In order to access the accessories port while maintaining the serial CCDI communication with the radio, a calibration box was used. The calibration box is a piece of hardware used internally by Tait Electronics Ltd to connect to the radio's internal circuitry via the accessories port. The calibration box is placed inside of the acoustically isolated environment, along with the speaker being tested and the other equipment as in the standard test setup. The only change in the physical connections of the test setup is that the serial cable connecting the computer and the receiving radio now passes through the calibration box, which also connects the radio and the speaker under test, as shown in Figure 18.

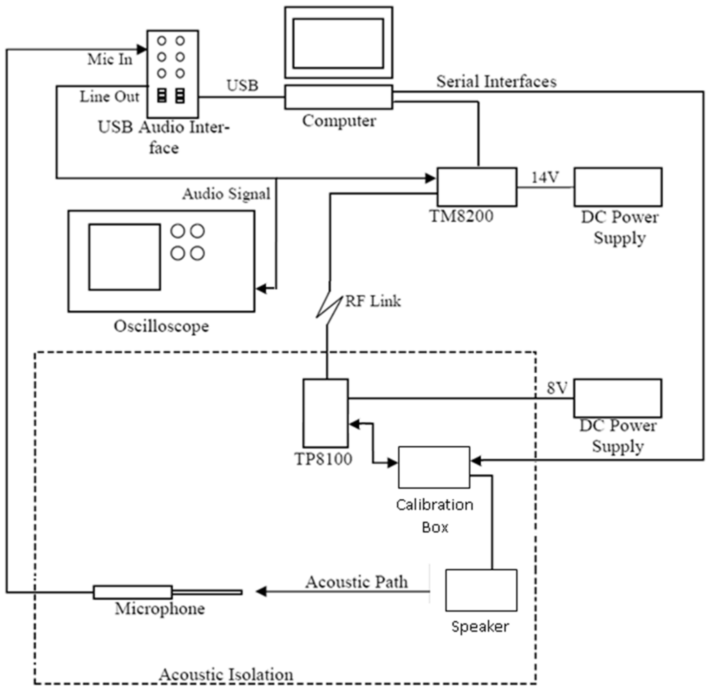


Figure 18: Physical test setup for testing speakers being driven by the radio's amplifier.

For the testing where the speaker was driven by a comparatively wide band, high power amplifier, the radios are omitted from the test setup. Although this means that the test setup no longer emulates real operating conditions, it does however provide information with regards to the speaker quality without the added distortions and disturbances that come about through RF transmissions. For this test setup the audio output of the USB interface was connected directly to the large amplifier, the output of which was connected to the speaker under test, as shown in Figure 19. Due to the removal of the radios there is no CCDI control in the system in this case, so the volume level was controlled by adjusting the output signal amplitude from the computer. This was done by the test suite program by adding a scaling factor to the audio output.

Different values for the scaling factor were used during the test effectively changing the volume level.

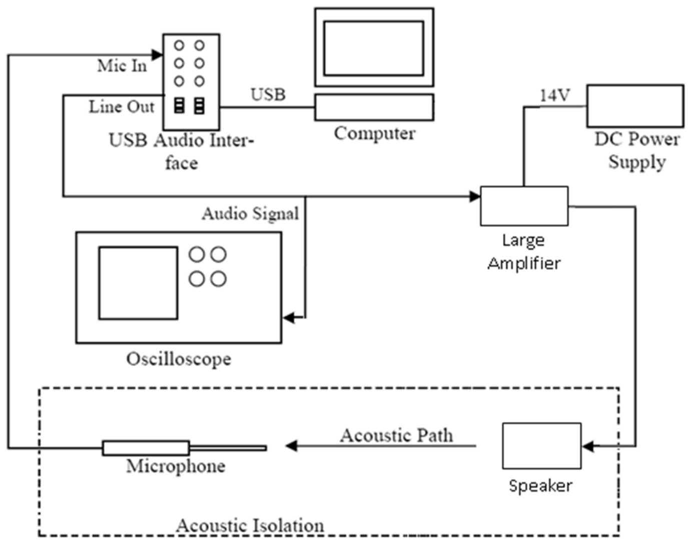


Figure 19: Physical test setup for speaker testing using a large amplifier.

The speaker testing used the described variations of the test suite to test the speaker's audio quality with respect to volume when being driven both by the radios and by a comparatively large, wideband amplifier. The boxes used to house the speakers for this testing provide a realistic and repeatable method for mounting the speakers both in a sealed and free field environment.

3.5.2 Radios without CCDI

In order to test radios without CCDI another variation of the test suite was developed. By testing radios which do not have CCDI the range of radios which can be tested was extended beyond Tait Electronics Ltd products. Comparisons can then be made with other products. The main problem with removing the CCDI interface from the test setup is that the ability to accurately reproduce tests is lost. This is because it is not a trivial matter to accurately position a variable volume control, such as a variable potentiometer, at any position between 100% and 0%. CCDI gets around this by setting the volume digitally, which means the same volume level can be set again at a later date. Also by removing CCDI the automated nature of the testing is lost with the volume level needing to be adjusted manually.

In order to validate the test results, instead of relating audio quality to a discrete volume level as in the normal test setup, the result is related to the Sound Pressure Level (SPL) of the radio when a volume level is set and white noise is played. There are also a number of weighting curves available which adjust SPL values to represent the perceived loudness of a signal. For this testing A weighting [27] was used, as the SPL range that the tests were performed in make this the optimal weighting curve. This means that although the volumes levels between tests may be different, the results are still comparable when plotted with respective SPLs.

4 Results

The results presented in this chapter represent a selection of results from tests that were performed throughout the duration of the project. The results are based around the testing of a number of different radios, as well as a variety of speakers. The analysis and testing methods used for this testing are described in the sections 2 and 3. The results provide an interesting insight into the effects that changes to the radios throughout the development process have had on the audio quality.

4.1 Indices of Speech Quality

4.1.1 Speaker Testing

Three speakers were assessed using the PESQ analysis method and speaker test suite setup described in sections 2 and 3. Speaker 1 was a high quality speaker, speaker 2 a copy of this high quality speaker and speaker 3 a cheaper speaker of a different design. Speakers 1 and 2 both had an impedance of 16 ohms with speaker 3 having an impedance of 8 ohms. All three speakers had a diameter of 35mm and were rated for 1 Watt rms.

One of the objectives of this testing was to determine the suitability of one of the speakers to replace the current speaker in an existing radio, with no other hardware changes. For this reason, during the radio driven tests no attempt was made to compensate for impedance differences between the three speakers. For the wide band amplifier tests, allowances were made for the difference in impedance. This was done by normalizing the volume levels to a set SPL, 100dB(A).

The impedance of the speaker currently in the radio is 8 ohms and thus the radio's internal amplifier was not over driven during any of these tests as all the speakers have equal or higher impedances. The fact that speakers 1 & 2 had a higher impedance then speaker 3 most likely caused a reduction in the power the speaker receives, which potentially had a detrimental effect on the maximum SPL produced by the speakers. This loss in SPL could however be regained by the speaker having a higher efficiency or sensitivity.

The results from this testing are broken up into two main groups, those using the wide band amplifier to drive the speakers and those using the radio.

Radio Driven Testing

The objective of these tests was to determine the performance of the three different speakers when they were driven in conditions similar to those of normal operation when mounted in a handheld radio. The PESQ score was measured at each of the radio's DSP volume setting, 0 through 25 for each speaker, with 25 being the loudest. This was initially done for the speakers in sealed enclosures, similar to the acoustic environment presented to the speakers currently used in the radios.

The results showed that speaker 3 had a better PESQ score at high volume levels then the other two speakers, as shown in Figure 20. All three speakers exhibited a sharp reduction in PESQ above volume level 16, with speaker 1 having the highest score. All three speakers suffered a sharp reduction in PESQ score, with the reduction from speaker 1 and 2 being much more severe. All three speakers then showed an increase in PESQ score at around volume level 22. Both of these observations are characteristic of the radio being used to drive the speakers as at this volume setting a parametric equalizer is activated which changes the frequency response of the audio path. This equalizer was originally implemented in order to reduce distortion at the higher volume levels.

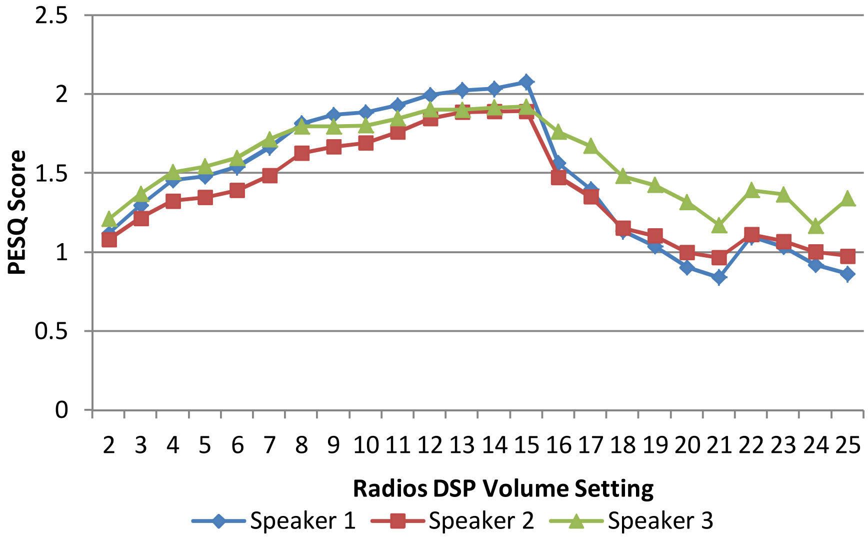


Figure 20: PESQ score for 3 speakers mounted in sealed enclosures and driven by radio 1.

As mentioned in the description of the speakers being tested, the speakers do not all have the same impedance and therefore there was the potential for speakers 1 and 2 to suffer a reduction in maximum SPL. In order to determine the effects of this difference in impedance a white noise test signal was played through the test suite and the A weighted SPL [27] was measured at the position of the microphone in the previous PESQ testing. The SPL was measured at each of the discrete volume settings of the radios DSP. The results showed speaker 3, the speaker with the lowest impedance, was generally louder than the other two, especially at the lower volume settings. At higher volume levels the difference became much smaller. The difference in volume at the lower volume levels between the 8 and 16 Ohm speakers was around 3dB, which seems reasonable as the 16 Ohm speakers will get half the power of the 8 Ohm speaker, as shown in equation (15).

$$P = \frac{V^2}{R} \qquad \textbf{(15)}$$

where P is power, V is voltage and R is the impedance.

The convergence of the two SPLs at the high volume level was most likely caused by a combination of speaker 1 and speaker 2 being able to produce louder SPLs, as the following wide band amplifier testing suggests, combined with the effects of the additional distortion products present at the higher volume levels. As the SPL convergence begins after volume level 21 the distortion products are once again likely characteristics of the radio being used to drive the speakers.

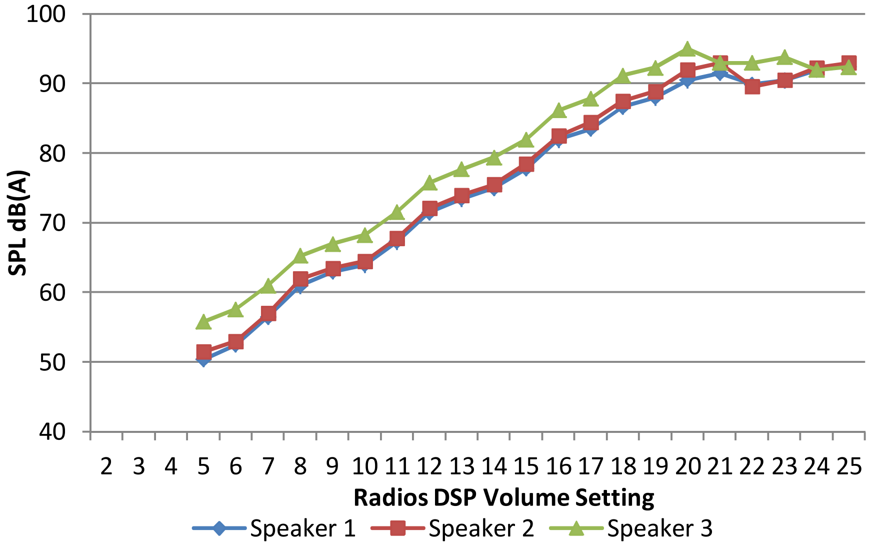


Figure 21 : SPL for 3 speakers mounted in sealed enclosures and driven by radio 1 with respect to radio 1's volume settings.

In order to determine the effect that running the speakers in a sealed enclosure had on the PESQ score, the PESQ and SPL test reported earlier in this section were repeated with the back of the enclosures removed in order to expose the speakers to a free field environment. The results showed a similar trend to the sealed enclosure testing, as shown in Figure 22. The notable differences between the two tests were the overall increase in PESQ score at the higher volume levels and the dramatic increase in rate of reduction of the score for speaker 3 between volume levels 15 and 16.

By putting the speakers in a free field environment, the speaker cone has less resistance to displacement then when a sealed environment is used. This means the magnitude of the cone's excursions increases which in turn causes an increase in SPL. This can be seen by the sealed and free field SPL measurements in Figure 21 and Figure 23. This higher excursion affects all speakers differently, depending on the stiffness of the speaker's spring and the excursion limits of the individual speakers. The change in the relationship between the three speaker's PESQ scores at the high volume levels (the reduction in difference between the scores) is likely to be caused by speakers 1 and 2 being able to handle the higher excursion better then speaker 3.

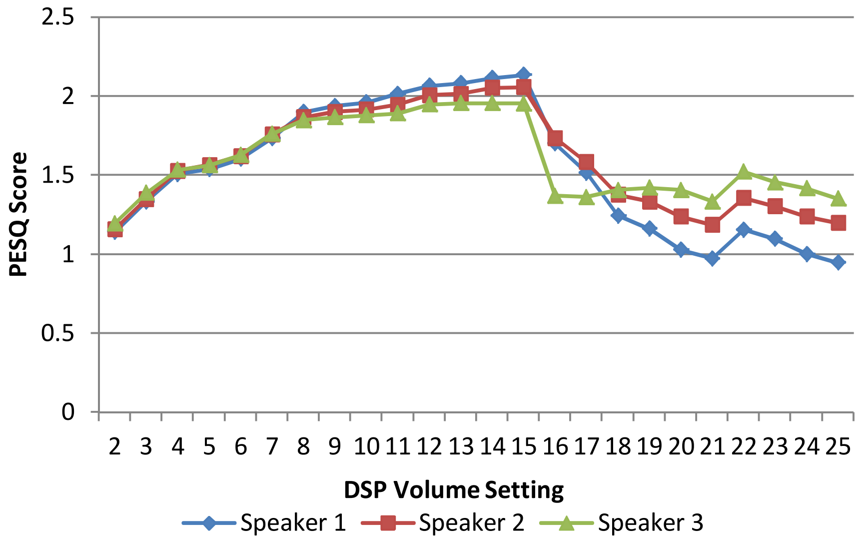


Figure 22: PESQ scores for 3 speakers in free space and driven by radio 1 with respect to radio 1's DSP volume settings.

Once again the A weighted SPL of each of the speakers was measured at each of the radio's volume levels with white noise being used as the test signal. For the free field testing each speaker's SPL stayed at a near constant distance from one another across the entire volume range, as opposed to the relationship seen in the sealed enclosure testing above. This result suggests that the merging of the three speakers' SPLs in the sealed enclosure test, Figure 21, is due to the speakers responding to the resonant properties of their enclosures. It is worth noting the reduction in SPL for all radios at volume level 21 is, as previously mentioned, a feature of the radio being used.

At the maximum volume setting, speaker 3 suffered a significant reduction in SPL, which the other two speakers did not replicate. This could either be a result of the speaker becoming overloaded and distorting heavily, due to lower impedance then the other speakers or simply a mis-calculation. Due to the fact that speaker 3 produced a higher SPL at volume level 21 than at the maximum volume position, an erroneous measurement is considered the most likely cause of the reduction.

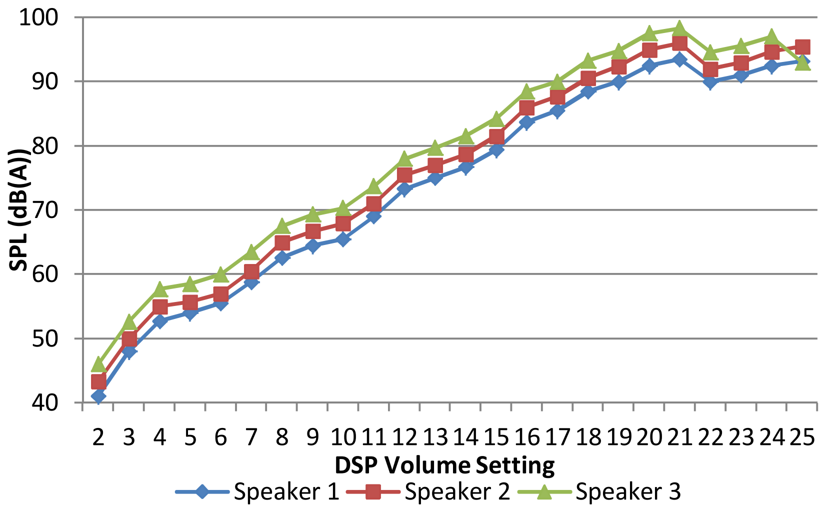


Figure 23:A weighted SPL for 3 speakers mounted in free space and driven by a TP8100 Release 1 radio with respect to the radio's DSP volume settings.

Wideband Amplifier Driven

After characterizing the different speakers whilst being driven by a radio, a comparatively wide band, high power amplifier was used to test the characteristics of the speakers without the added distortion generated by the radios. This test provided an insight into how the speakers compared to one another in more ideal conditions then the radio's amplifier output.

In order to compensate for the impedance mismatch, instead of simply swapping the speakers out with one another and keeping the amplification level the same, the amplifier was adjusted so that each speaker produced 100dB(A) when a full amplitude white noise signal was played. This method has the advantage of characterizing each of the speakers' audio quality at the same SPL level. Unfortunately, speaker 3 was not able to achieve the 100dB(A) target while in a sealed enclosure and in an attempt at free field testing, the speaker was destroyed before the required SPL was met.

Once the amplifier's volume control had been set, the test signals were played by the computer with sequentially reduced amplitudes. At each of the iterations the PESQ result was measured for each speaker. The results showed that speaker 1 had a higher PESQ score across all of the volume levels, with a trend fairly closely matched by speaker 2. Speaker 3 had the lowest PESQ score with a greater variation being shown with respect to volume level, as shown in Figure 24. This was most likely because speaker 3 was struggling to produce the 100 dB SPL target and was therefore at the top of its range of operation. If the volume level was increased for the other two speakers it was expected that a similar effect would occur. This was not verified however because 100 dB was already considered loud enough for this testing and on increase may have resulted in the destruction of the other two speakers.

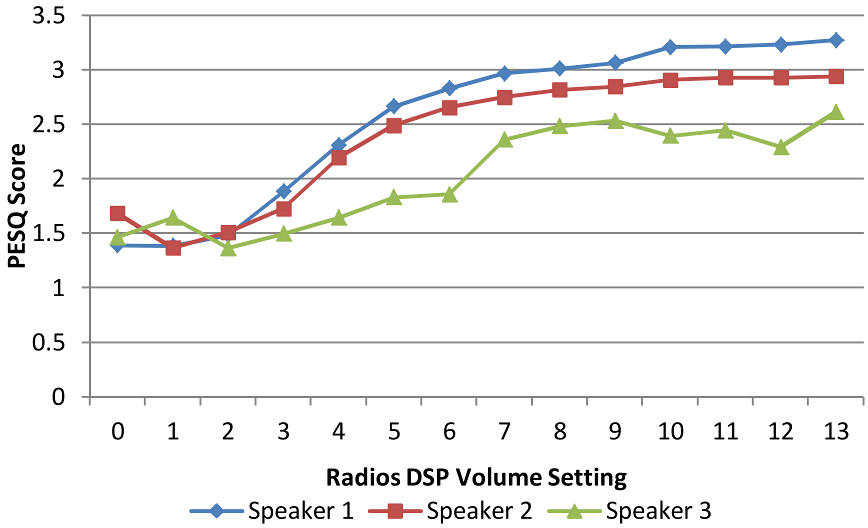


Figure 24: PESQ score for three speakers mounted in sealed enclosures and driven by a large, wideband amplifier with respect to increasing volume level.

The same A weighted SPL measurement that was performed in the radio driven speaker testing was conducted for each of the sequential volume settings with the three speakers connected to the wide band amplifier. The results showed that speakers 1 and 2 had reasonably identical SPL levels across the volume range and speaker 3 was near a constant 7 dB less, as shown in Figure 25. After volume setting 13, the SPL of speaker 3 appeared to reach a maximum level, whereas the other two speakers continued to increase. This supports the assumption that speaker 3 was near its maximum volume limit and thus causing the PESQ score to be lower and more variable than the other two speakers.

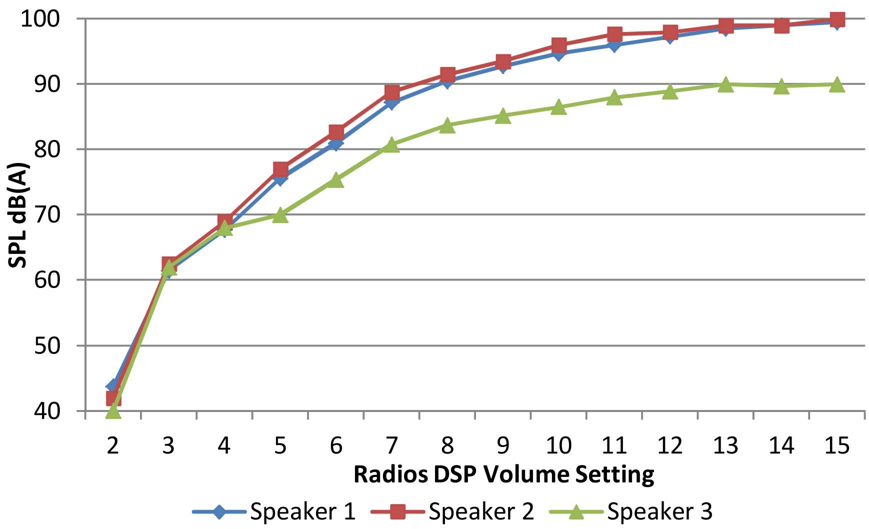


Figure 25: A weighted SPL for 3 speakers mounted in sealed enclosures and driven by a large, wide band amplifier with respect to increasing volume level.

The SPL's measured for speaker 1 in the radio driven tests exceeded the maximum of 90 dB(A) measured in the wide band amplifier tests. This is due to frequency weightings of the A weighted SPL calculation. A weighting is most sensitive at 2.5 kHz as the human ear is most sensitive that [27]. The radio being used for the radio driven testing also has a resonate peak in its output spectrum at 1 kHz, which has the first harmonic distortion product at around 2 kHz. This means that the white noise test signal used for the SPL measurements is shaped to have a large spectral peak near the more sensitive regions of the A weighting system for the radio driven tests, whereas for the wide band amplifier tests this resonant peak was not present. The power going through the speaker for the wide band testing was therefore more evenly distributed across the frequency spectrum; resulting in a lower A weighted SPL reading.

4.1.2 *Radio Testing*

The radios that were tested represent three radios from the same series at different stages of development. Each radio was operating in narrow band mode in the H5 band, 464.575 MHz. Radio 1 represented the earliest form of the series, which is characterised by having a parametric equaliser that was activated at around volume level 21 up to level 25 in a single step. Radio 2 had the same equaliser that was activated over the same volume range but in a smoother fashion. Radio 3 performed the same processing on the audio line as radio 2 but had a new, more powerful class D amplifier.

Perceptual Evaluation of Speech Quality (PESQ)

As was the case with the speaker testing, the PESQ result was calculated at each of the radio's volume settings. The results showed Radio 1 to have an increase in PESQ score at volume level 21, which was where the parametric equaliser was activated, as shown in Figure 26. The abruptness of the change in PESQ score differentiates radio 1 from radio 2. The results showed radio 3 to be the better radio, in a PESQ score sense, until after volume level 19, at which point radio 2 becomes the better performer.

These results are somewhat misleading. Although the volume setting is discrete and absolute in a DSP sense, they do not necessarily represent the same SPL output from the radios. This means that although radio 2 appeared to be producing a higher PESQ score, it may not be reaching the same volume levels that the other two radios are.

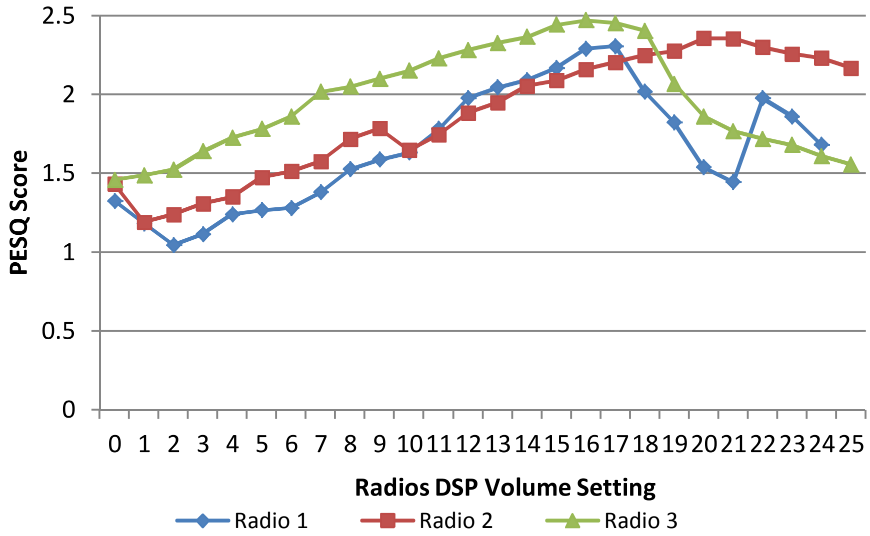


Figure 26: PESQ result with respect to volume level for radios 1 - 3.

In order to verify this, the PESQ results from the previous test needed to be normalised to the SPL level, effectively transforming the results into the SPL domain. The first step in this process was to measure the SPL level produced by the radios at each of the discrete volume settings, using full scale white noise as a test signal. The A weighted SPL was measured at the position of the microphone in the anechoic chamber (Section 3.2). The results from this test revealed that at the high volume levels, radio 3 was the loudest radio, followed by and matched up to volume level 21, by radio 1. Radio 2 was shown to be the quieter radio by as much as 3dB around volume level 21, as shown in Figure 27.

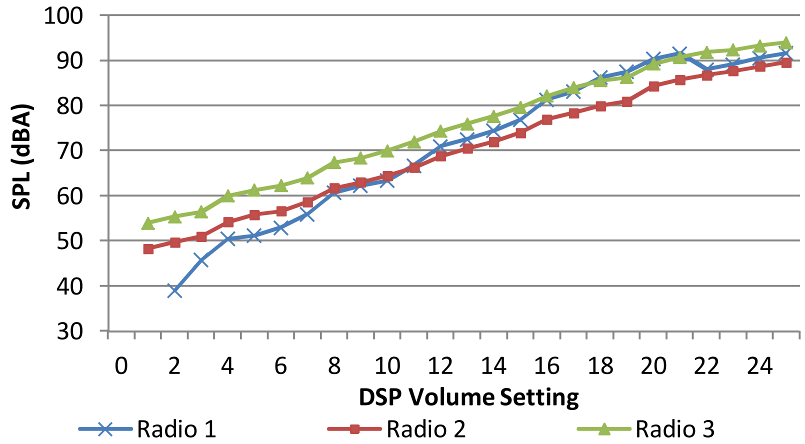


Figure 27: Relationship between A weighted SPL and DSP volume setting for the 3 radios.

Using the results from this measurement, the PESQ score was then plotted with respect to the A weighted SPL produced by the radio at each of its discrete volume settings. This showed that, although radio 2 had the highest PESQ score at the top volume setting, it is also quieter than the other two radios. This is an undesirable characteristic as users demand high SPL levels from their radios as well as optimum audio quality.

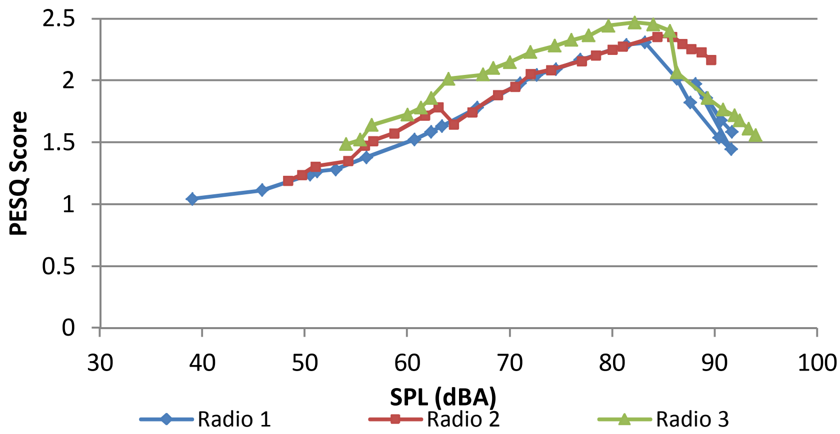


Figure 28: PESQ score with respect to SPL (dB(A)) for radio's 1-3.

The results shown in Figure 28 also demonstrated that at the lower volume levels, radio 3 had a higher score then the other two radios, which also have a similar PESQ to one another.

Radio 3's new class D amplifier is the likely cause of the higher PESQ score with respect to volume level, compared to the other two radios.

Upon completion of these tests it was noted that the PESQ standard [1] recommended that the test signals be made up of multiple 1-3 second speech bursts. The signals used for the previous testing consisted of single 1-3 second speech bursts. Investigations were conducted in order to determine the effect of test signal length on the PESQ result. The first noticeable difference between the two result groups was that there was a general increase in PESQ value for the longer test signals. Radio 1 no longer surpassed radio 3's quality at volume level 22 and radio's 1 and 3 were also now closer in PESQ score to radio 2 at the high volume levels, as shown in Figure 29.

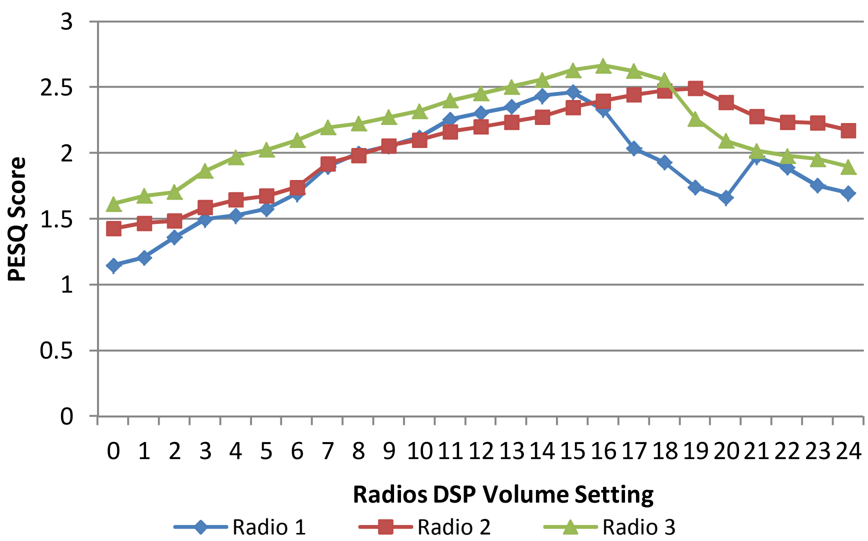


Figure 29: PESQ score with respect to each radios DSP volume setting for radio's 1-3 with long test signals.

Once again these results were normalised with SPL for each of the radios, the result of which is displayed in Figure 30. These results showed the radios to have similar PESQ scores at the lower volume levels. Radio 1 had a sharp reduction in PESQ score starting at SPLs above

80 dB, with radio 2 producing higher SPLs before suffering the same reduction at a lesser rate. Radio 3's PESQ score shared a similar trend to radio 2, but at a higher SPL and ultimately a higher PESQ score. This was expected as the difference between the two radios, in an audio sense, was the inclusion of a higher power amplifier in radio 3. This meant that more power was produced from the amplifier with less additive distortion.

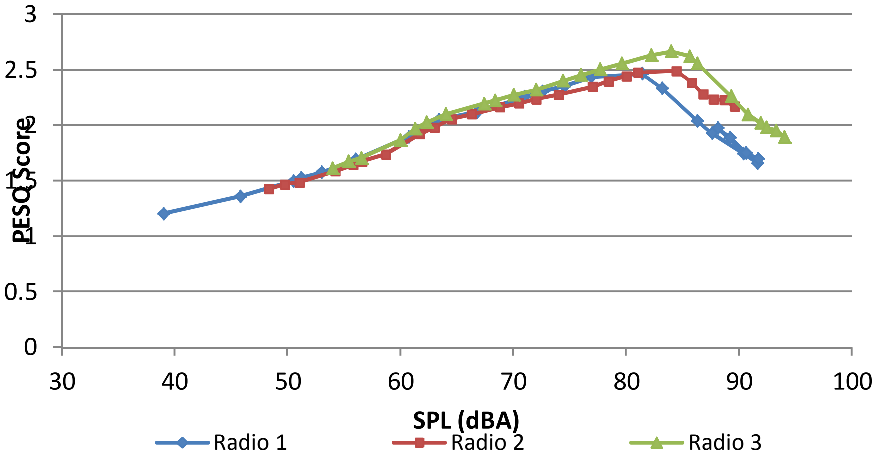


Figure 30: PESQ score with respect to SPL(A) for radio's 1-3 using long test signals.

Speech Transmission Index (STI)

The STI is a measure of intelligibility as described section 2.2. The result is a number between 0 and 1, 1 representing high intelligibility and 0 representing low intelligibility. The STI was also used to analyze the three radios described above with respect to discrete volume levels.

The results showed a similar trend to those shown in the PESQ results, with radio 2 outperforming the other two radios at the top volume levels and radio 1 exhibiting the lowest STI, as shown in Figure 31. A gradual increase to a peak in the intelligibility is followed by a reduction, most likely due to added distortion for both radios 1 and 3.

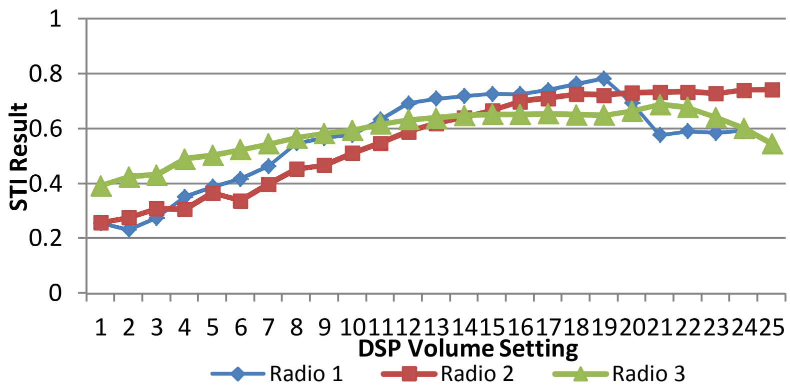


Figure 31: STI result with respect to DSP Volume level for radio's 1 – 3

SPL normalization was performed on the STI result. Radio 2 had the highest STI of the three radios, but it was also quieter than the other two radios, as shown in Figure 32. The results also showed a sharp degradation in intelligibility for both radio 1 and radio 3 at the top volume levels. This was likely to be caused by the high level of distortion present at those levels due to clipping of the amplifier and speaker.

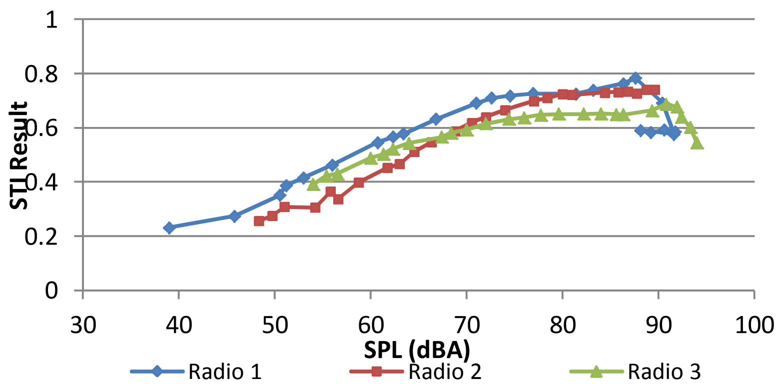


Figure 32: STI results with respect to SPL(dB(A)) for radio's 1-3.

4.2 Vowel Space Analysis

As part of this testing, the formant frequencies of a number of the early test signals were hand picked in order to verify the accuracy of the results obtained from the automated formant extraction process. The hand picking test was conducted during the early stages of this project and at that time, radio 3 was not available. Early on in the planning stages of this research it was desided that an objective method of measuring formant frequencies would be used. An automatic formant extraction system was choosen and implemented. In order to validate the results produced from this system a number of the formant frequencies were manually extracted using a predetermined criteria.

4.2.1 Hand Picked Formants

The first formant frequencies determined were those of the original sythesized vowel sounds, prior to any degradation. The formant frequencies of the 3 vowels were determined by visual inspection and use of the LPC spectrum, calculated using the method described in section 2.3.2. Once the extraction process was completed, formant 1 was plotted with respect to formant 2 for each of the vowel sounds, as shown in Figure 33, and the distribution of the vowel sounds was examined.

Using the formant frequencies found by the manual extraction method, the vowel space was calculated using the equations defined in section 2.3.2. The vowel space of the original vowel sounds prior to any degradation was calculated to be 137268 Hz^2.

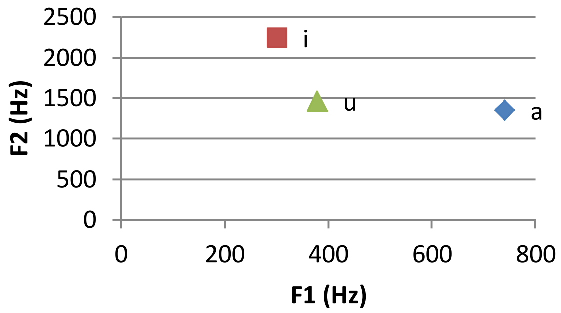


Figure 33: Handpicked vowel space of original synthesised vowels.

The next stage of testing required the vowel sounds to be played through the two radios using the test suite described in section 3. For each radio the vowels were recorded at all of the radio's volume settings. Once the recorded signals had been obtained the formant frequencies were extracted using the manual formant extraction system described earlier in this section. The outcome of this procedure was the first and second formant frequencies of the three corner vowels, after being received through the radios, with respect to the radios volume settings. This is shown in Figure 34 and Figure 35

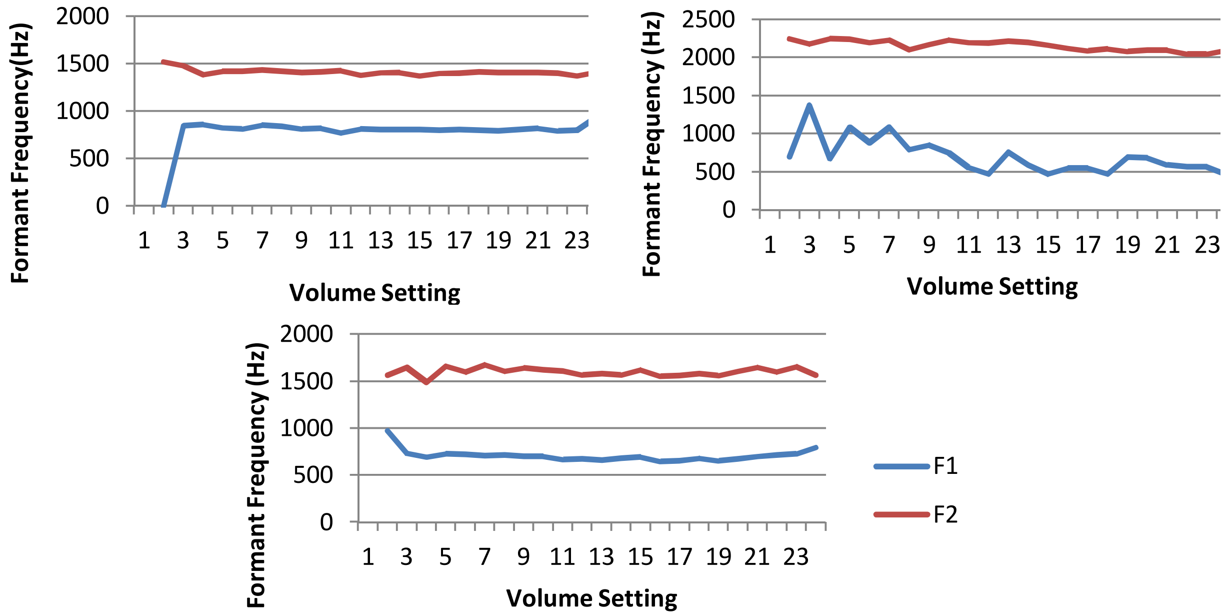


Figure 34: Handpicked formant 1 and 2 with respect to volume level for radio 1 for the vowels a (Top Left), i (Top Right) and u (Bottom).

The effect of the radio transmission on the formant frequencies appears to be relatively consistent across all volume levels for radio 1, particularly for the /a/ and /u/ sounds. Formant 1 of the /i/ sound shows more variation in value as the volume is increased. Both /a/ and /u/ show formant 1 increasing in value at or near the top volume setting of the radio.

A similar increase in formant 1 frequency of the /a/ and /u/ vowel sounds was seen in the results for radio 2. Overall the results appear to be similar to one another, with radio 2 showing more variability in formant 1 frequencies for the /u/ vowel sound and less variability in formant 1 frequencies for the /i/ sound.

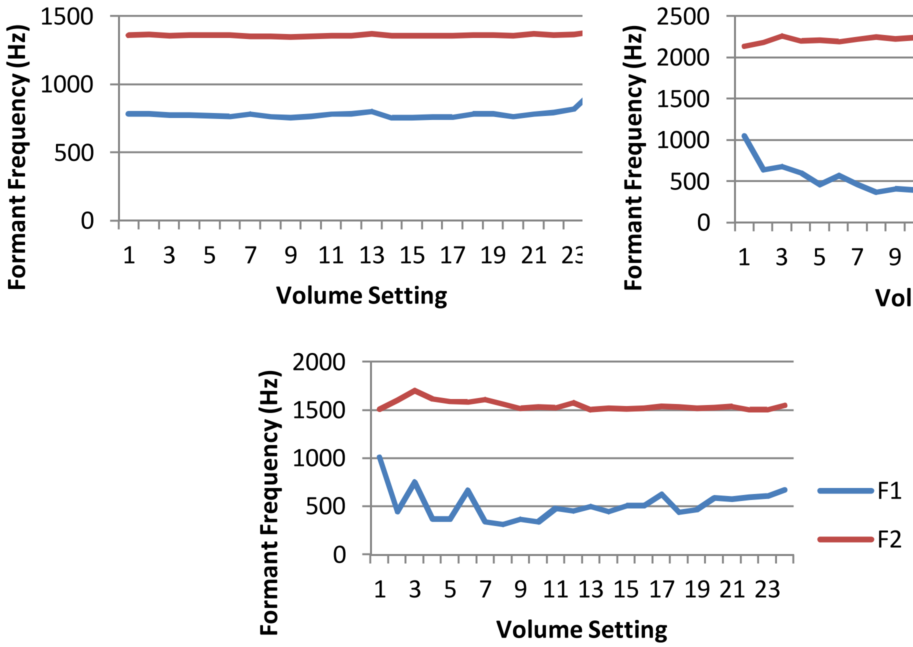


Figure 35: Handpicked formant 1 and 2 with respect to volume level for radio 2 for the vowels /a/(Top Left), /i/ (Top Right) and /u/ (Bottom).

The formant frequencies were plotted with respect to one another, meaning the effect of the radio transmission on the shape of the vowel space could be examined for a selection of volume levels, as shown in Figure 36 and Figure 37.

This showed that although the formant frequencies appeared consistent in Figure 34, the shape of the vowel space in the F1 – F2 plane was greatly affected. For radio 1 at the lower volume levels (15) the location of the /u/ vowel sound had shifted from having a formant 1 frequency of around 400 Hz, up to near 800 Hz. The distribution of the vowel sounds changed dramatically both with respect to volume level and when compared to the original vowel sounds.

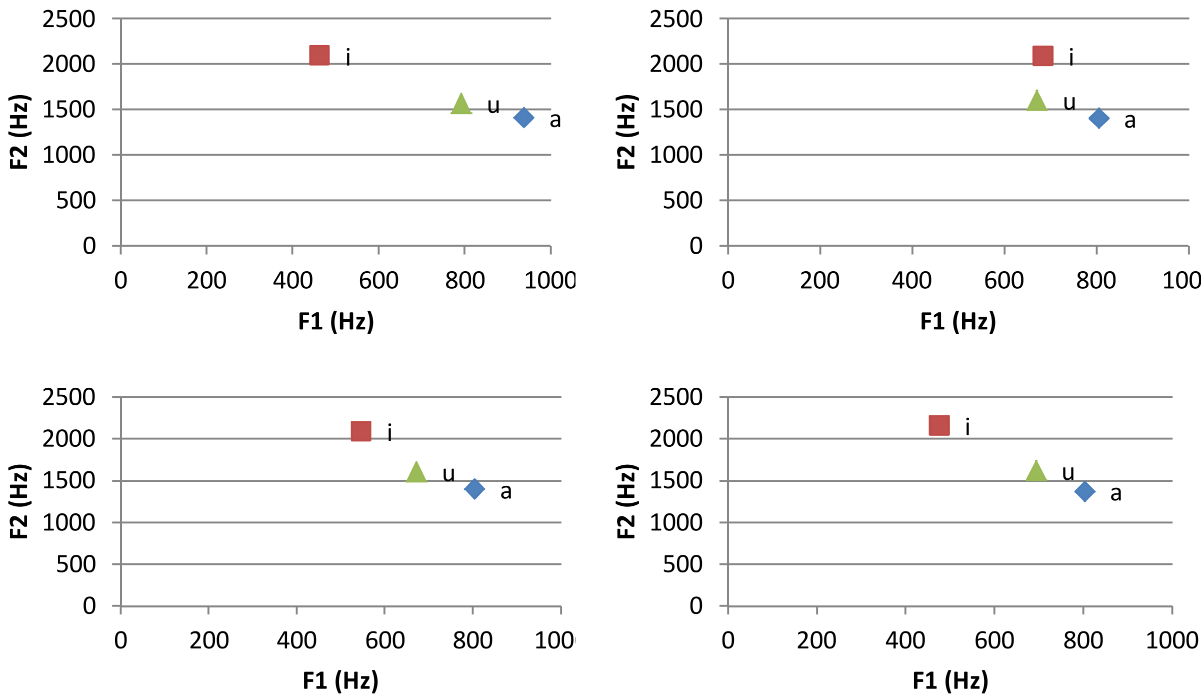


Figure 36: Vowel space of radio 1 using handpicked formants at volume levels 24(Top Left), 20 (Top Right), 17 (Bottom Left) and 15 (Bottom Right).

The vowel space generated by the three vowel sounds for radio 2 showed a similar reduction to radio 1. The shift in vowel space shape for radio 2 did not appear as severe as that observed for radio 1. At volume level 15 the vowel space still resembled the distribution of the vowels seen for the original vowel sounds (c.f Figure 33), as well as at volume level 24. Both the /i/ and /u/ vowel sounds appeared to have suffered a substantial increase in formant 1 frequency, as seen for radio 1.

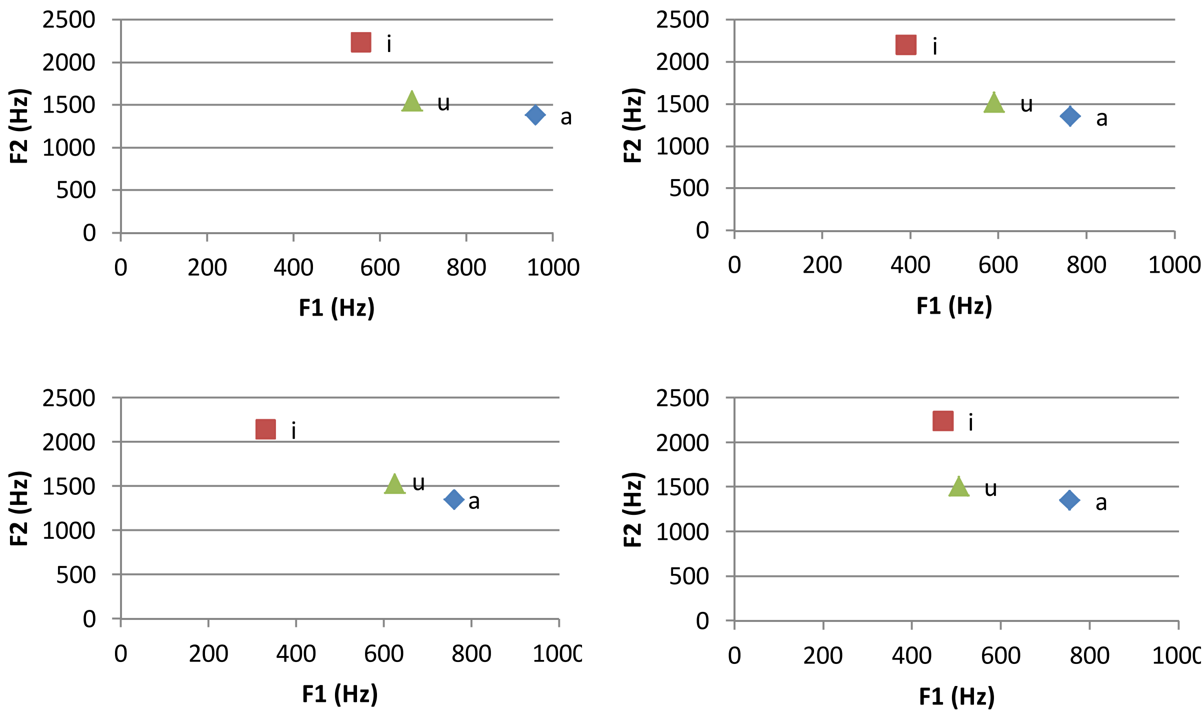


Figure 37: Vowel space of radio 2 using handpicked formants at volume levels 24(Top Left), 20 (Top Right), 17 (Bottom Left) and 15 (Bottom Right).

Using the formant frequencies determined by the hand picking method, the vowel space was calculated with respect to volume level for each of the radios, as shown in Figure 38. This showed that the vowel space had a large amount of variability across the volume range, revealing an inverse relationship with increasing volume level. Radio 1 was shown to have an overall smaller vowel space than that produced by radio 2, with radio 2 also having a less variable result through the middle volume levels.

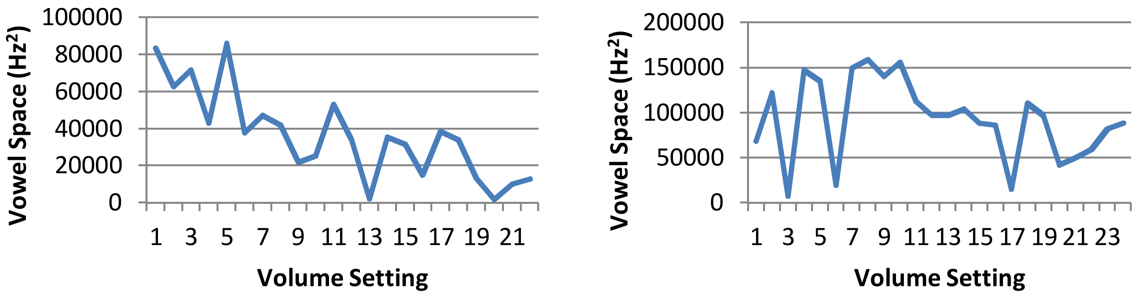


Figure 38: Vowel space with respect to volume level for radio 1 (left) and radio 2 (right) using handpicked formant frequencies.

The results from the hand picking of the formant frequencies revealed that radio 2 had a larger vowel space than radio 1, though both radios substantially reduced the vowel space of the original test signals. The area of the vowel space measured for radio 2 showed that the arrangment of the three vowels in the F1-F2 plane is a better representation of the original vowel space than radio 1. The need for an automated method was highlighted by the number of test signals used and the time required for manual formant extraction.

4.2.2 Automated Formant Extraction.

The automated formant extraction process described in section 2.3.2 was used in conjunction with the synthesized vowel sounds in order to determine the effect of a radio transmission on the first two formant frequencies of the /a/, /i/ and /u/ vowel sounds. The test was conducted on each of the three radios, as radio 3 had now become available for testing. After this the calculated formant frequencies for each of the vowel sounds were used to determine the vowel space of each radio with respect to volume.

The formant frequencies and vowel space were calculated for the original test files using the automated formant extraction method. This was used as a benchmark for determining the effect that the radios had on the vowel space and also to compare the results between the automated extraction method and the manual system.

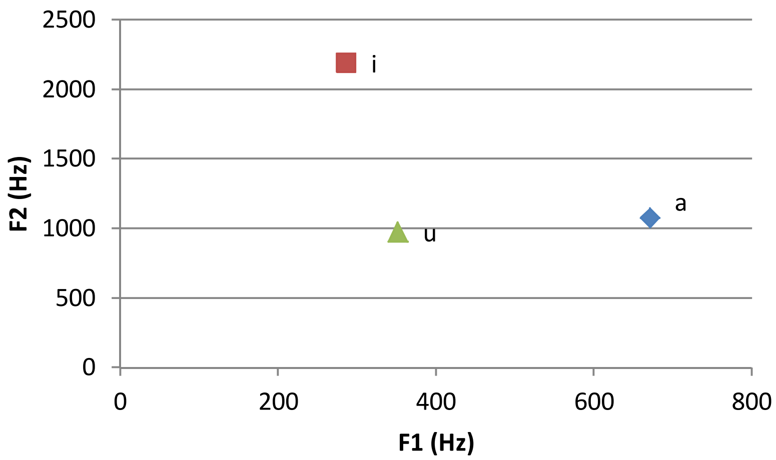


Figure 39: Vowel space of the original synthesised vowel sounds via automatic formant extraction.

The results showed that the formant frequencies were close to those from the hand picked results, as shown in Figure 39, with the exception of the 2nd formant values for the /a/ and /u/ vowels being around 400 Hz lower for automated extraction then for handpicked formants. The vowel space was then calculated using the formant frequencies and the equation described in section 2.3.2. This gave the vowel space of the original vowel sounds to be 197044 Hz^2, which is larger than the area calculated from the hand picked formant frequencies, mainly due to the lower formant 2 freqencies in the automated tests.

Radios 1,2, and 3 were each tested using the test setup described in section 3, where the three vowel sounds were played through each radio and the output signal recorded. The formant frequencies were determined for each of the recorded vowel sounds using the automated formant extraction method. This procedure generated the first and second formant frequency for the three corner vowels with respect to volume level for the three radios, as shown in Figure 40, 41 and 42.

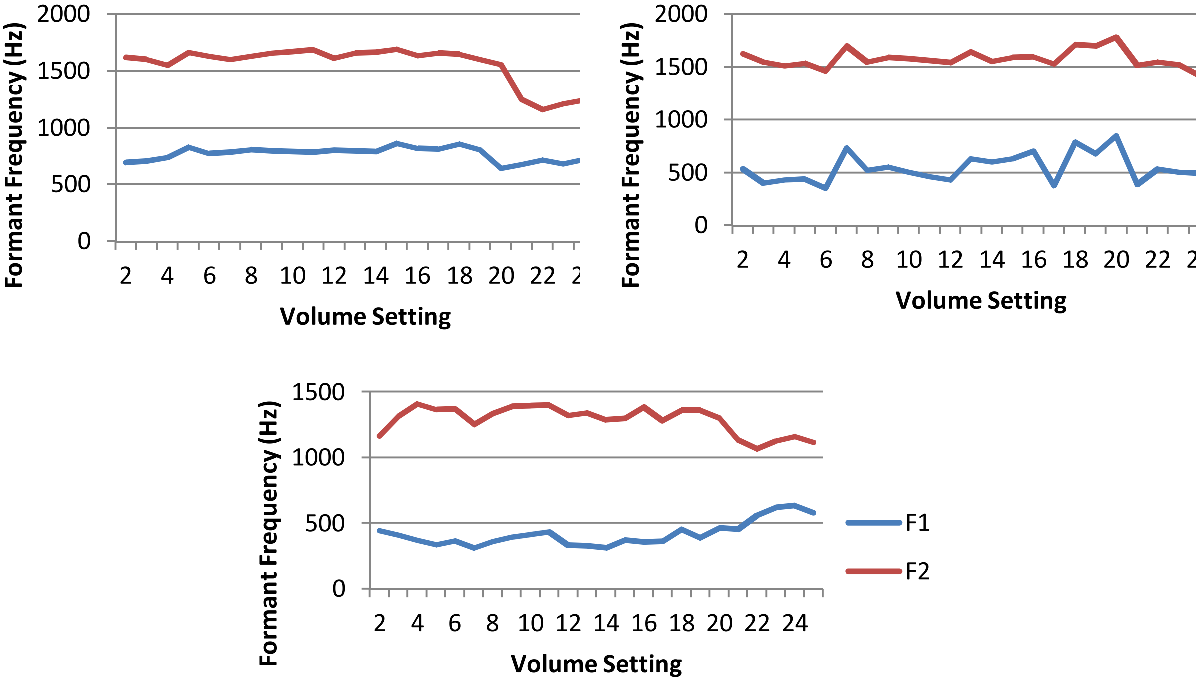


Figure 40: Formant 1 and 2 with respect to volume level for radio 1 for the vowels /a/ (Top Left), /i/ (Top Right) and /u/ (Bottom).

The formant frequencies showed a similar level of consistency across the volume levels to the hand picked tests. A transition in formant frequency at the higher volume levels was noted for both formant 1 and 2 for the /a/ and the /u/ vowel sounds for radio 1. This transition also began at a lower volume setting than the hand picked testing showed. The /i/ vowel sound still showed some variability in formant 1 value accross the volume range, but of a smaller magnitude than was seen in the handpicked formant tests.

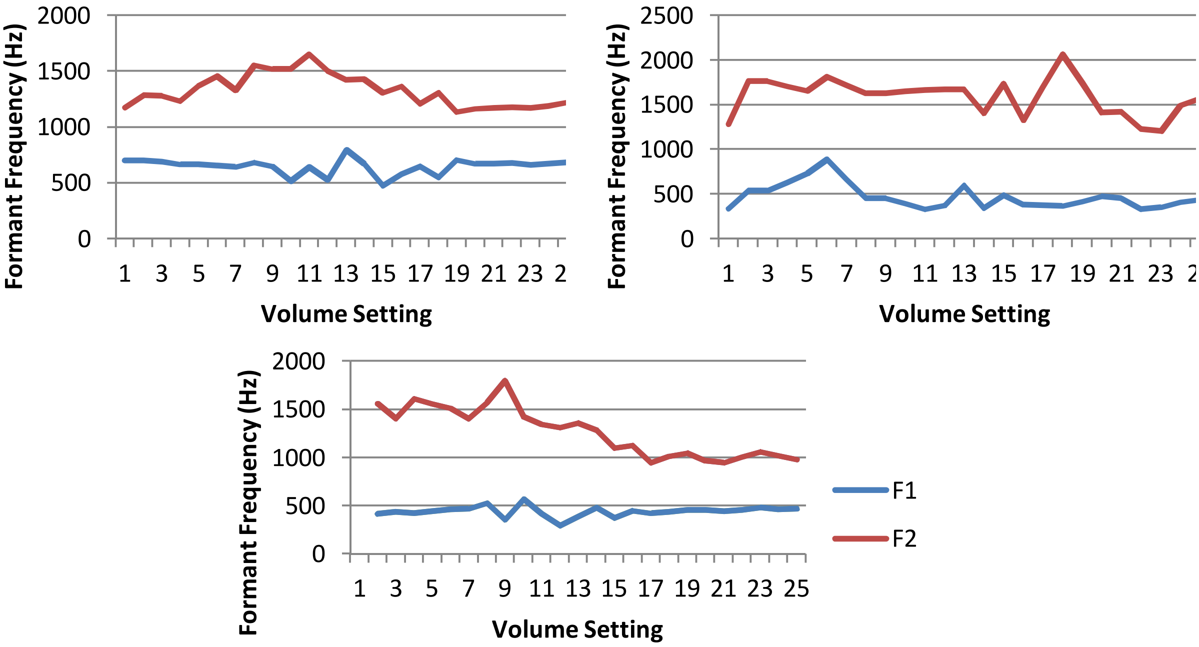


Figure 41: Formant 1 and 2 with respect to volume level for radio 2 for the vowels /a/(Top Left), /i/ (Top Right) and /u/ (Bottom).

The results from analysis of the signals produced by radio 2 showed that the first formant frequency values of the /a/ vowel did not reduce at the high volume levels as they did for radio 1. The second formant frequency began to reduce at a lower volume level but reached a similar value to the final second formant of radio 1. The /u/ vowel began to reduce in formant 2 value much early then radio 1, with the final value being below that of radio 1's /u/ vowel second formant value.

Radio 3 was also tested using the same method, though there were no hand measured results to compare to the automated measurements. The results showed that the formant frequencies of the /u/ vowel sound were similar to those seen for radio 2, with a similar level of fluctuation and an inverse relationship between frequency and increasing volume level. The formant frequency values of the /a/ vowel were the most consistant across the volume range out of all three of the radios. The /i/ vowel showed an inverse relationship between formant frequency and increasing volume for both formant 1 and 2, with the results becoming more consistent at the higher volume levels.

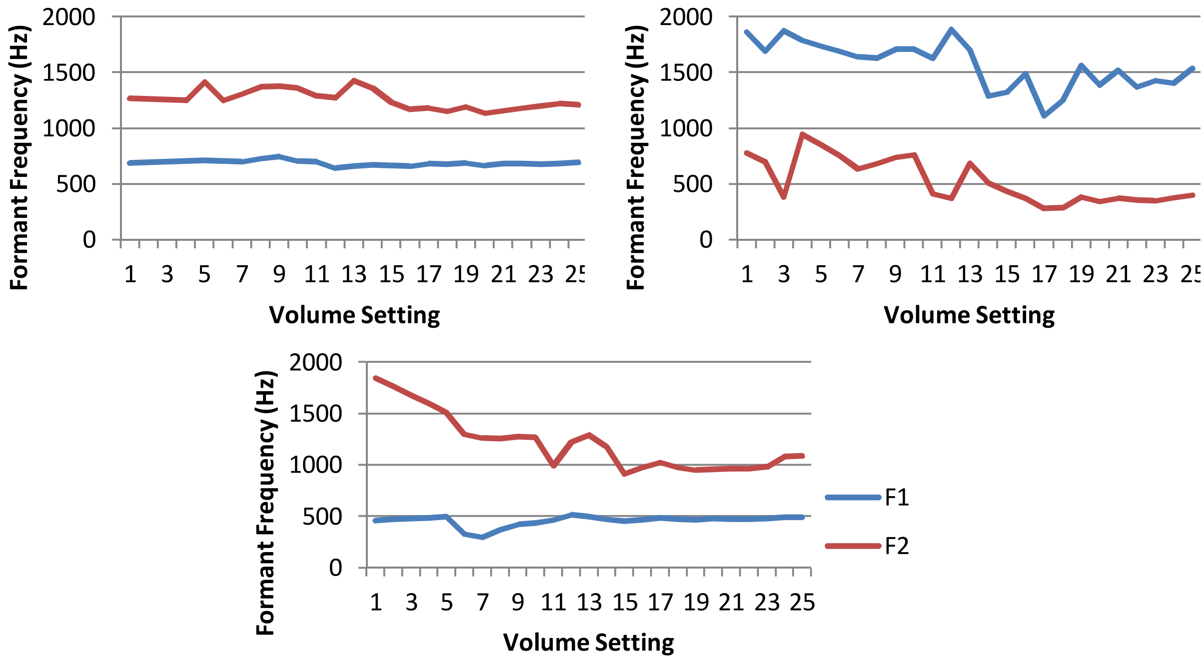


Figure 42: Formant 1 and 2 with respect to volume level for radio 3 for the vowels /a/ (Top Left), /i/ (Top Right) and /u/ (Bottom).

The formant frequencies for each of the vowel sounds were plotted in the F1 - F2 plane for a selection of volume levels for each of the radios, as shown in Figure 43,44, and 45.

Analysing the radio 1 signals showed that at the higher volume levels the distibution of the vowel sounds in the F1 –F2 plane was similar to the original vowel sounds, compared to both the hand picked test results and lower volume levels.

The distribution of the vowel sounds in the F1 – F2 space after being played through radio 2 appeared to resemble the distribution of the original vowel sounds seen in Figure 39 more closely than those played by radio 1. The distribution of the vowel sounds with respect to one other across all of the volume levels was also more consistent for radio 2 then radio 1. The variation in the first formant frequency value for the /a/ vowel was less for radio 2 than it was for radio 1. The formant 2 frequency of the /i/ vowel, sound, however appeared to show more variation in the radio 2 testing compared to radio 1.

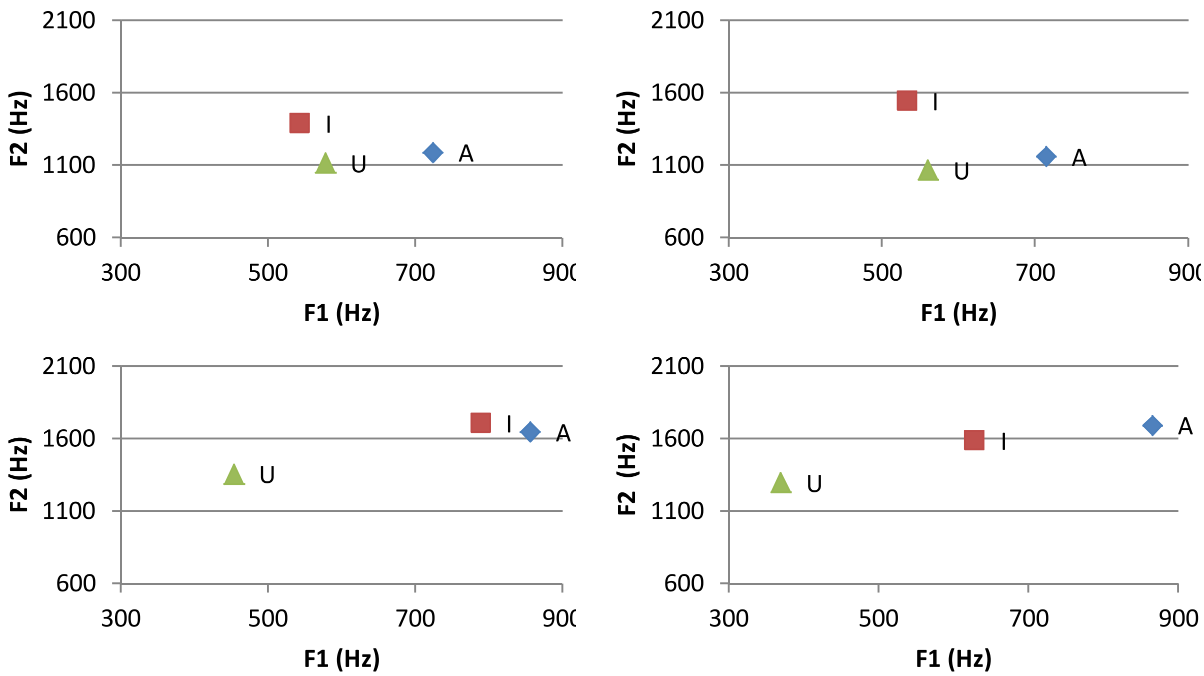


Figure 43: Vowel space of radio 1 at volume levels 25(Top Left), 22 (Top Right), 18 (Bottom Left) and 15 (Bottom Right).

The positioning of the three vowel sounds in the F1 –F2 plane for radio 3 revealed the most consistent distribution of vowel sounds of all three of the radios, with only small variations in the postion of the vowels with relation to one another, as shown in Figure 45. The distribution of the vowel sounds also remains similar to that of the original vowel sounds, as shown in Figure 39. The vowel with the most variability in its position is the /i/ vowel, though it always stays above and left of the other two vowels, which is not the case with the other two radios.

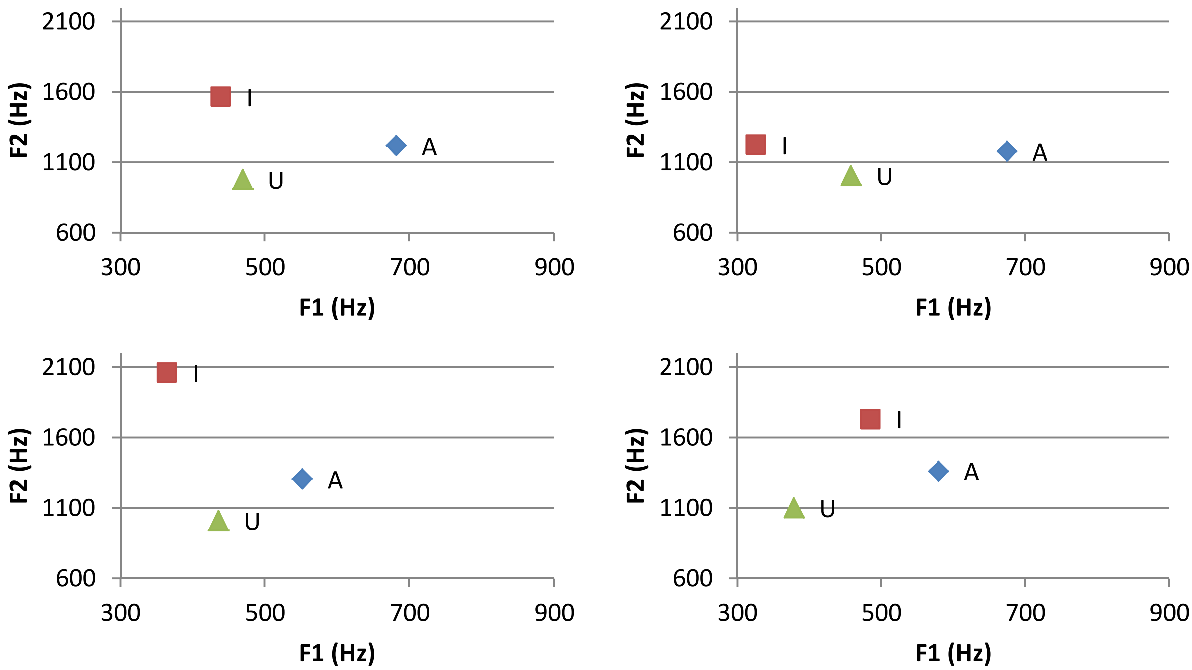


Figure 44: Vowel space of radio 2 at volume levels 25(Top Left), 22 (Top Right), 18 (Bottom Left) and 15 (Bottom Right).

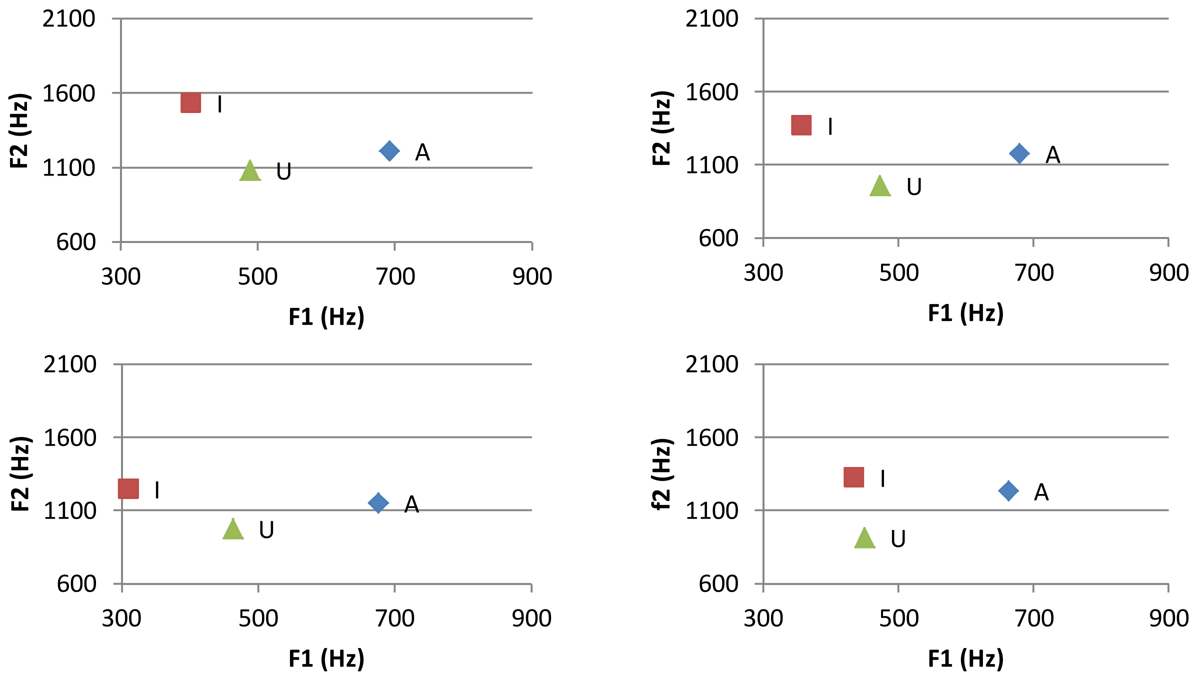


Figure 45: Vowel space of radio 3 at volume levels 25(Top Left), 22 (Top Right), 18 (Bottom Left) and 15 (Bottom Right).

The vowel space was calculated using the formant frequencies that were calculated using the automated formant extraction process and the vowel space equation described in section 2.3.2. The vowel space was determined for each test signal recorded at all the volume levels of the three radios, as shown in Figure 46, 47, and 48.

The vowel space calculated for all three of the radios showed a large amount of variablity with respect to volume level. Overall radio 3 was shown to have the largest vowel space at the very low volume levels. Radio 2 produced the largest vowel space at the medium volume levels (16 – 19) as well as at the maximum volume setting.

The vowel space calculation using the handpicked formant frequencies in section 4.2.1 revealed a inverse relationship between the vowel space and increasing volume level, this relationship is not evident in the results of the automated extraction process. The automated extraction results do show radio 1 to have the overall lowest vowel space values, which agrees with the results of the manual extraction process.

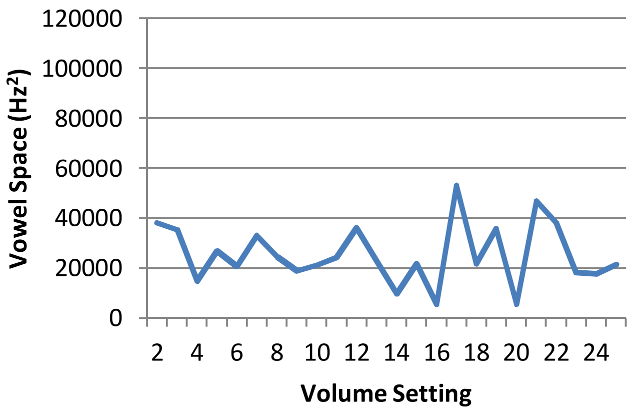


Figure 46: Vowel space with respect to volume for radio 1 using automated formant extraction.

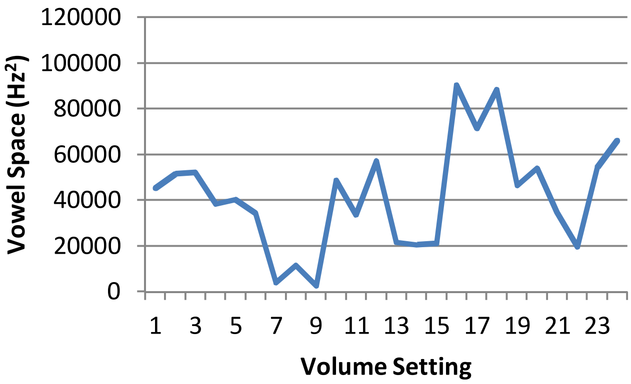


Figure 47: Vowel Space with respect to volume for radio 2 using automated formant extraction.

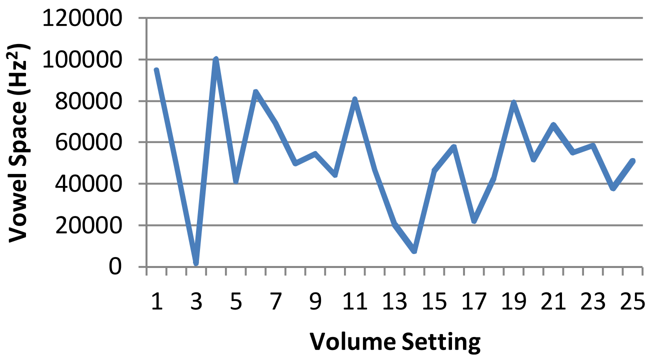


Figure 48: Vowel space with respect to volume for radio 3 using automated formant extraction.

The automated formant extraction results showed that radio 1 had the smallest vowel space of the three radios, with radio 2 and 3 having similar spaces to one another. The vowel space measure had a large amount of variability in results across the volume levels for all three radios. The results measured here did not correlate highly with those of the handpicked formant testing, although the same relationship between vowel space and radio1 and 2 was found. This could be due to changes in the recorded test signals as a result of variations in the test suite, which occurred as a part of its development throughout the duration of the project.

4.3 Phonetic Analysis

Using the vowel sounds that were recorded as part of the vowel space analysis, a number of other parameters were extracted to determine the effect of a radio transmission on perceptual level. The parameters were Singer Power Ratio and F1 F2 Slope, both of which are described in section 2.4. The following two sections present the results obtained from this analysis.

4.3.1 Singer Power Ratio

The Singer Power Ratio was calculated using the formants calculated from both the manual and automated extract methods. The Singer Power Ratio is another phonetic metric that is used for determining how well a person's voice carries over noise.

Manual Analysis

Using the formant frequencies and magnitudes extracted during manual vowel space analysis (section 4.2.1), the Singer Power Ratio was calculated for the original vowel sounds prior to any degradation, for radios 1 and 2 with respect to volume. The Singer Power Ratio of the original vowel sounds was calculated to be -8.2 dB, 1.5 dB and -8.2 dB for the /a/, /i/ and /u/ vowel sounds respectively.

The Singer Power Ratio for the three vowel sounds when played via radios 1 and 2 with increasing volume is shown in Figure 49. When comparing the singer power ratio of the signals produced by the radios to the Singer Power Ratio of the original signal, the only similar result was the /a/ vowel sound when produced by radio 2. The calculated Singer Power Ratio of the other configurations appeared to be much larger than that of the original result. Overall, the Singer Power Ratio did not appear to be larger for any one radio across all of the vowel sounds.

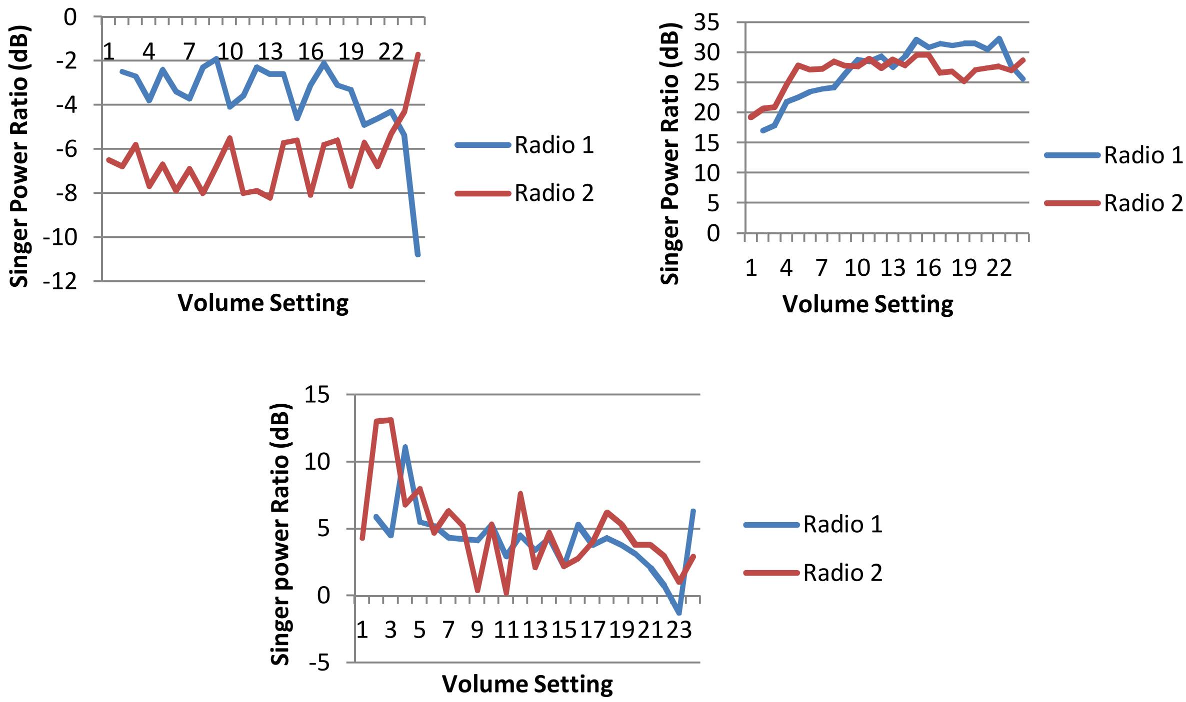


Figure 49: Singer Power Ratio for radio 1, with respect to volume setting, for the vowels A(Top Left), I (Top Right) and U (Bottom).

Automated Analysis

The Singer Power Ratio was then determined using the automated formant extraction results, with the inclusion of radio 3. Using this method the original /a/, /i/ and /u/ vowel sounds were calculated to have Singer Power Ratio values of 42.1 dB, 45.7dB and 49.5 dB respectively.

The Singer Power Ratio of radios 1, 2 and 3 was then calculated with respect to volume level for the /a/, /i/ and /u/ vowel sounds, shown in Figure 50, 51 and 52. The values that were calculated were dramatically different to those from the manual formant extract results. The

Singer Power Ratio for all three of the radios and all three of the vowel sounds were much closer to the values that were calculated for the original signals.

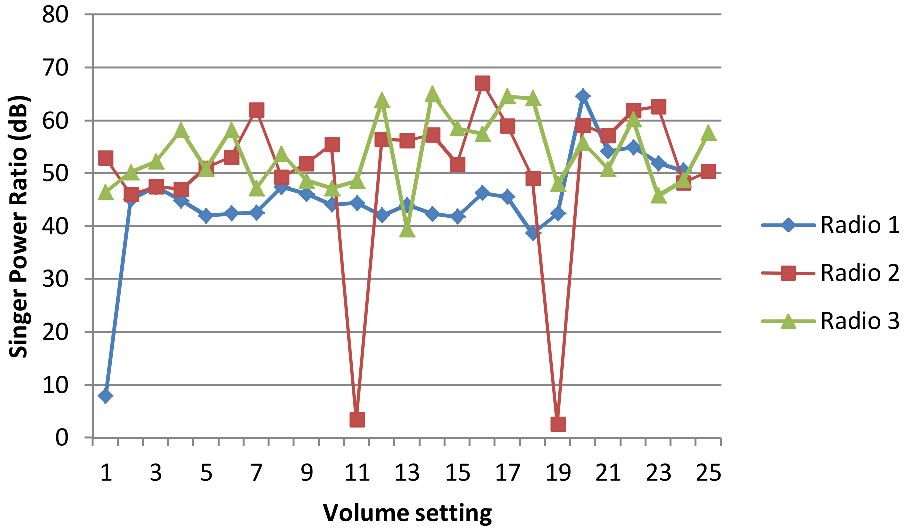


Figure 50: Singer Power Ratio of A vowel sound with respect to volume level for radios 1-3.

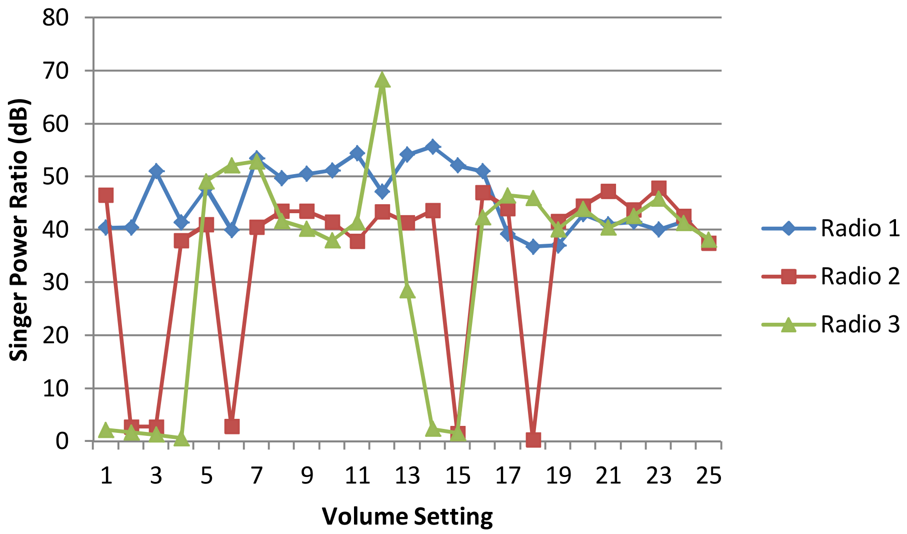


Figure 51:Singer Power Ratio of I vowel sound with respect to volume level for radios 1-3.

The Singer Power Ratio of radios 2 and 3 appeared to be relatively similar to one another across the volume range for each of the three vowel sounds. Radio 1 was shown to have a Singer Power Ratio similar to the other two radios at the higher volume levels for the /a/ and /i/

vowel sounds and lower for /u/.At the mid range volume levels (11-20) radio 1 had a smaller singer power ratio for the /a/ and /u/ vowel sounds compared to the other 2 radios.

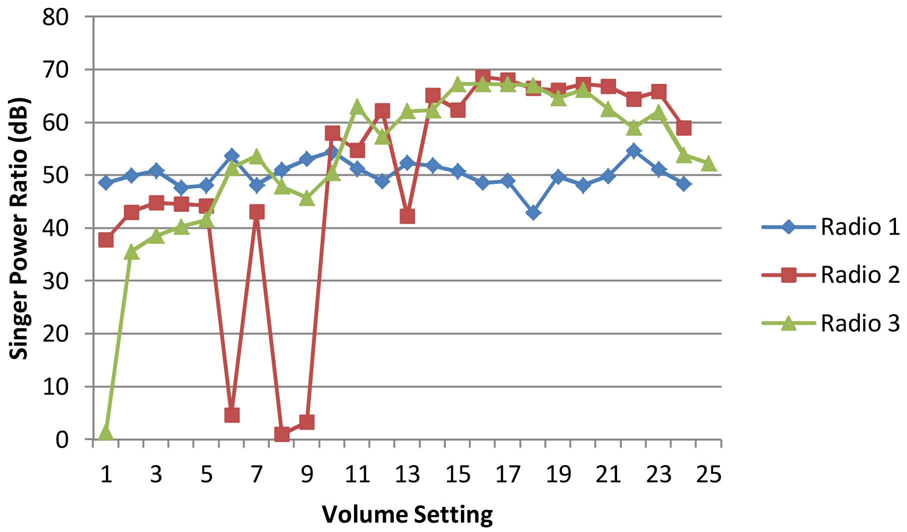


Figure 52: Singer power ratio of U vowel sound with respect to volume level for radios 1-3.

Overall radio 1 was consistently the closest to the original Singer Power Ratios of the original signals. At the higher volume levels of the /a/ vowel sound, radios 2 and 3 had a similar result to radio 1. Comparatively, for the /i/ vowel sound, radios 2 and 3 were the closest to the original value.

The results from the manual Singer Power Ratio analysis of the three radios were significantly different from the values determined for the original vowel sounds. This was not the case for the automated test analysis method, with original and degraded results both having similar values. Assuming that a minimal change in Singer Power Ratio between original and degraded signal is optimal, this testing does not correlate with the PESQ and STI results from the previous section, with both tests suggesting that radio 1 is the better performer. From the automated tests, radios 2 and 3 showed very similar characteristics across the entire volume range which is not what the PESQ and STI results suggest, due to the high amount of distortion present in radio 3 at the higher volume levels.

4.3.2 F1 F2 Slope

The F1 F2 slope was calculated using the formant frequencies and magnitudes that were determined by the automated formant extraction process used for the vowel space testing. Using this method the F1 F2 slope of the original three vowel sounds was deemed to be -0.035, 0.0051, -0.083 for the /a/, /i/ and /u/ vowel sounds respectively.

The F1 F2 slope was then determined with respect to volume for the three vowel sounds, after being played through radios 1,2 and 3, as shown in Figure 53, 54 and 55.

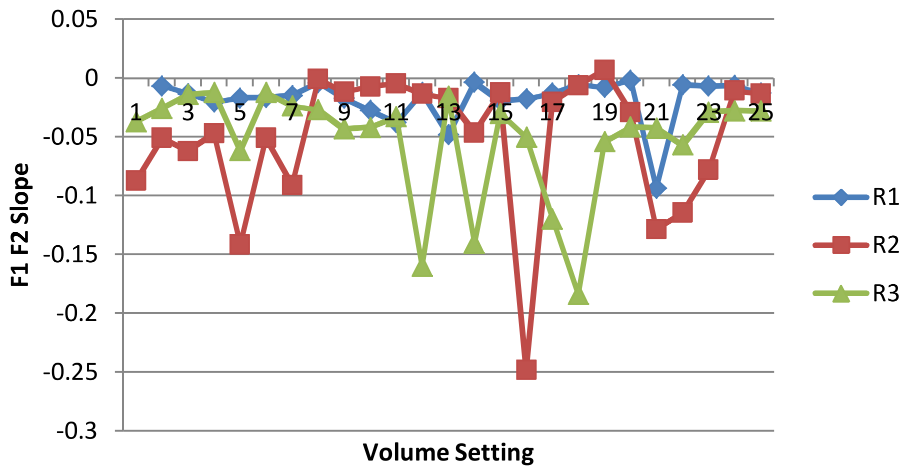


Figure 53: F1 F2 slope of the A vowel with respect to volume for radios 1-3.

For all of the vowel sounds, radio 1 was shown to have a F1 F2 slope close to 0, with small amount of variation in the lower volume levels of the /i/ vowel sound. The other 2 radios also appeared to also have a F1 F2 slope close to 0, except for the /u/ vowel sound, where they both dropped to a value of -0.4 at the upper mid range volume levels. The F1 F2 slope of the /i/

vowel sound calculated for radio 2 showed a spike around volume level 19, which is not matched by radio 3.

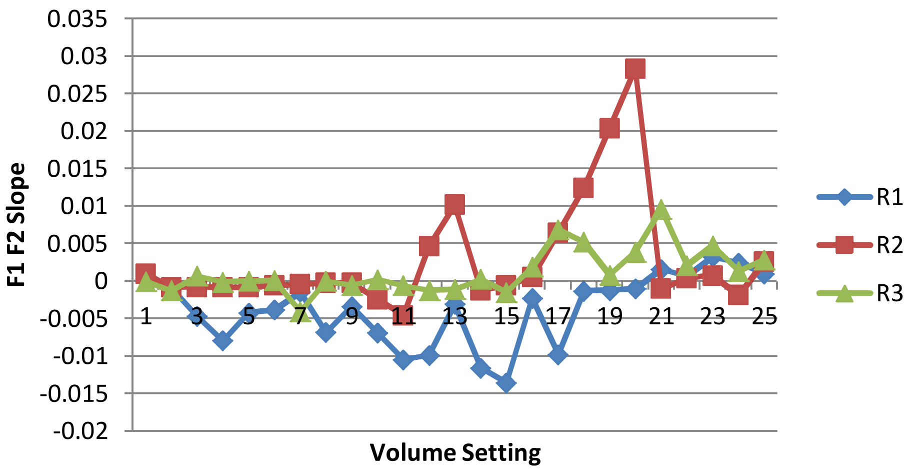


Figure 54: F1 F2 slope of the I vowel with respect to volume for radios 1-3.

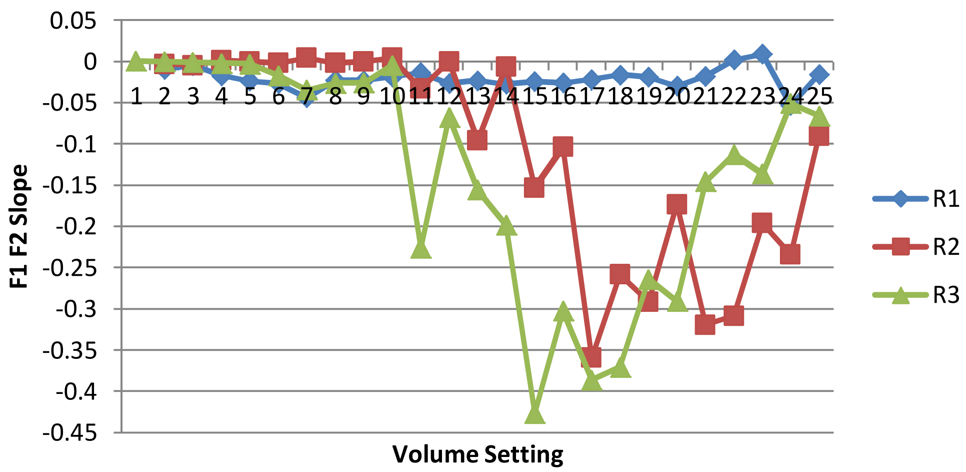


Figure 55: F1 F2 slope of the U vowel with respect to volume for radios 1-3.

Overall there was a reasonable amount of variation in F1 F2 slope with respect to volume level for all three of the vowel sounds. The significant deviations from a F1 F2 slope of 0 did not appear to correlate with any results from other analysis methods. As was the case with the

Singer Power Ratio, the F1 F2 result did not show a high correlation with the results produced by the PESQ and STI tests.

4.4 Frequency Response Testing

The frequency response was measured using the method described in section 2.5 for each of the three radios at their maximum volume setting. Due to the band limited nature of the radios it was assumed that the frequency content above 4 kHz was a product of the distortions generated after the DSP in the radio, by the amplifier and speaker, as the radio was running at a sampling rate of 8 kHz. These measurements gave information about the spectral distribution of the sound that the radios produced, not SPL, as the equipment used was not calibrated to provide this functionality.

Figure 56 shows the response of radio 1 and indicates a section of high density spectral energy in the frequency band between 300 Hz and 4000 Hz, which also defines the passband of the radio. The frequency content in the passband was not evenly distributed, with a number of troughs and peaks visible. There was no obvious peak visible at or around 1 kHz where the parametric filter should be active. This was likely to be a result of the addition of harmonic and inter-modular distortions masking the peak.

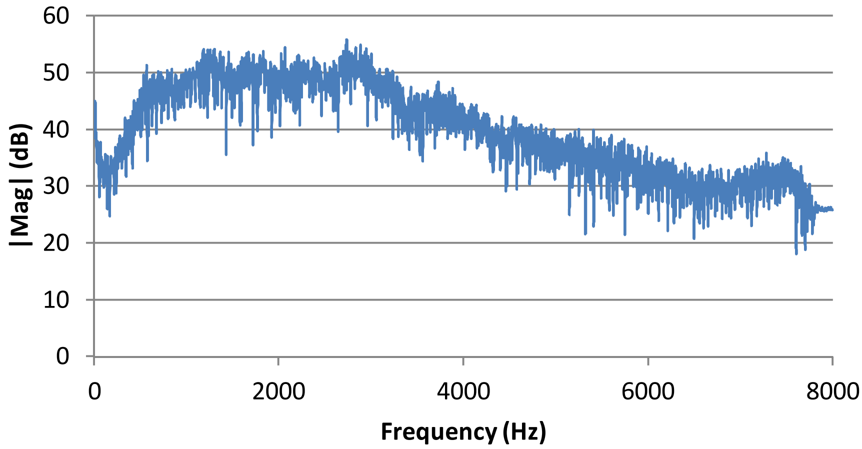


Figure 56: Measured frequency response of radio 1.

Radio 2 had a much more defined passband, with a steep drop off around 3700 Hz, as shown in Figure 57. This suggested that the radio was not producing as many distortion products as the other two radios, as the frequency content present appears to be concentrated within the passband of the radio. The response also had a peak visible around 1000 Hz which was likely to have been caused by the parametric filter that was activated at the higher volume levels. Overall the passband response of radio 2 was not particularly flat, although there appeared to be less distortion present than for the other two radios suggest.

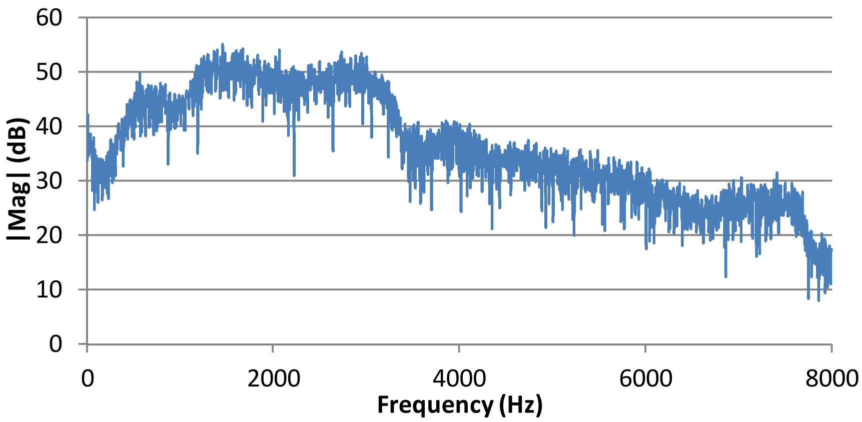


Figure 57: Measured frequency response of radio 2.

The frequency response measured for radio 3 was the flattest of the three radios, with a slight reduction in output around 3700 Hz. This suggested that radio 3 is producing more harmonic distortion products than radio 2 but less then radio 1.

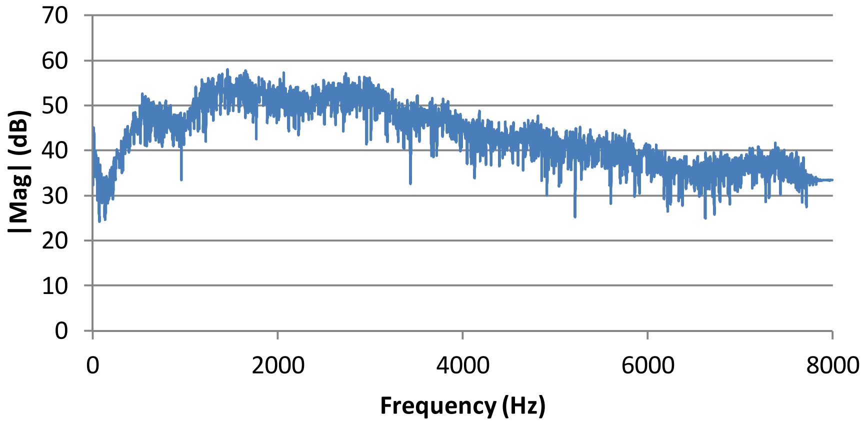


Figure 58: Measured frequency response of radio 3.

5 Validation Testing

A number of validation tests were carried out in order to determine the accuracy and robustness of the test procedures used, and give validity to the results produced by the audio quality test suite.

5.1 Method

The main method employed to check the validity of the tests was to systematically degrade the output signal of the radio with the addition of noise, then, using the analysis methods used and described in chapters 3 and 4, determine the effects on the audio quality. The expectation of this testing was to see an overall degradation in the audio quality as a result of the added noise.

In order to conduct these tests the same test setup and procedures were used as for the testing in section 4.1.2, with the addition of a comparatively high power, wide band loud speaker, as shown in Figure 59. The loud speaker was used for adding the noise signal into the free field in which the radio was located (anechoic environment). The noise signal used was white noise and was supplied by a mains powered signal generator. The speaker's amplifier was set to full volume and the SPL of the white noise was adjusted by changing the voltage amplitude of the output of the signal generator. This improved the repeatability of the experiment as any bandwidth variations caused by varying the amplifiers gain were mitigated due to the constant amplification setting. For this testing the white noise was set to an SPL of 70 dB(A), measured at the position of the microphone used for the recording the output of the speaker.

The test suite program was then run as normal, with white noise constantly being added, for each of the three radios being tested. The PESQ and STI measures were calculated for each of the discrete volume settings, as per the tests with results presented in section 4.1.2.

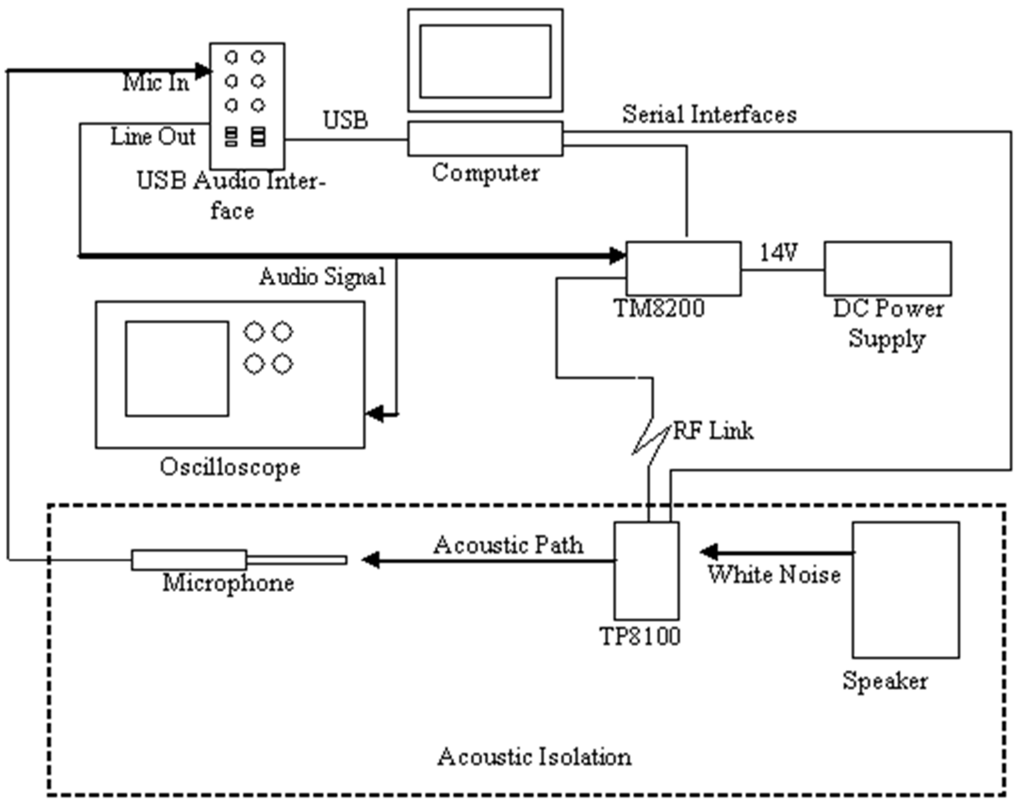


Figure 59: Physical test setup variation used for validation testing.

Another part of the validation testing was to test a radio from another series, measuring the PESQ score and STI with respect to volume and then compare the results with those obtained in the above section. The hypothesis of this investigation was that the results found here will have a similar shaped PESQ score, when plotted with respect to volume, to that of the other radios tested. The radio being tested is of a different design and operates in the B1 band, 151.625 MHz and for the purpose of this thesis is denoted ‘radio X’.

The test remains essentially the same for this test. The only changes are the replacement of the transmitting radio for one that transmits in the B band and the removal of the serial volume control. This means that the volume level is manually set for each of the test volume iterations. In order to provide a meaningful reference variable, the A weighted SPL the radio produces at each of these volume levels was measured from the position of the microphone when full scale white noise was being received.

The final test to be performed was to play and record the PESQ and STI test signals through a comparatively wideband, high power amplifier. Using the resulting recordings the PESQ score and STI were to be calculated. The expected result was that both analysis methods would reveal the loud speaker to have comparatively high quality and intelligibility in comparison to the radios.

5.2 Results

Noise testing

The first set of results obtained from this testing were the PESQ scores for radios 1, 2 and 3 across all of the volume levels with the addition of parasitic noise. The first thing to note with regards to these results was the large fluctuation in PESQ score at the lower volume levels, as shown in Figure 60. This is most likely due to the algorithm not detecting the degraded speech signal from the recorded output due to complete masking by the added noise and not a characteristic of the radios. It was decided to disregard these lower volume results and instead focus on the PESQ scores for the high volume levels, which are in any case of more interest.

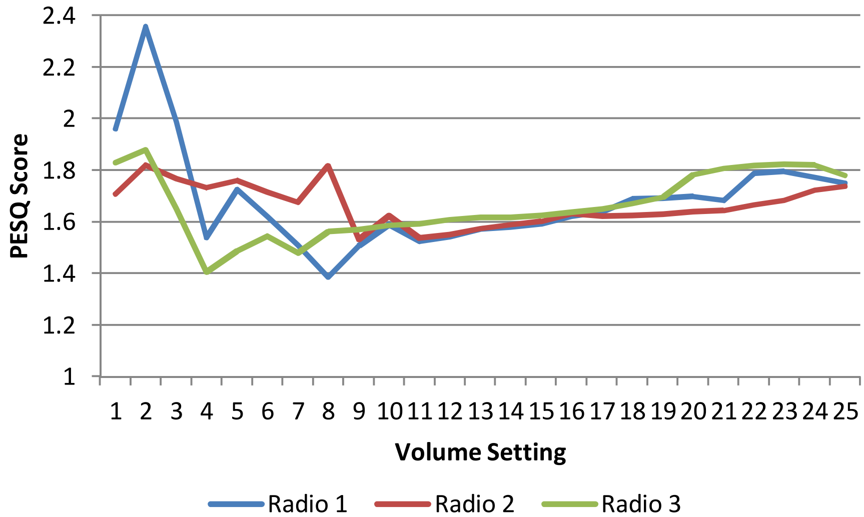


Figure 60:Results of the PESQ test of Radio 1,2 and 3 with parasitic white noise added.

By removing the lower volume levels from the result shown in Figure 60, the differences between the three radio's PESQ scores with the addition of noise become more apparent, as shown in Figure 61. The first thing to note is the net reduction in PESQ score with the addition of the noise signal. This can be seen by comparing the PESQ results from chapter 4 with those from this test, as shown in Figure 62 and Figure 61 respectively.

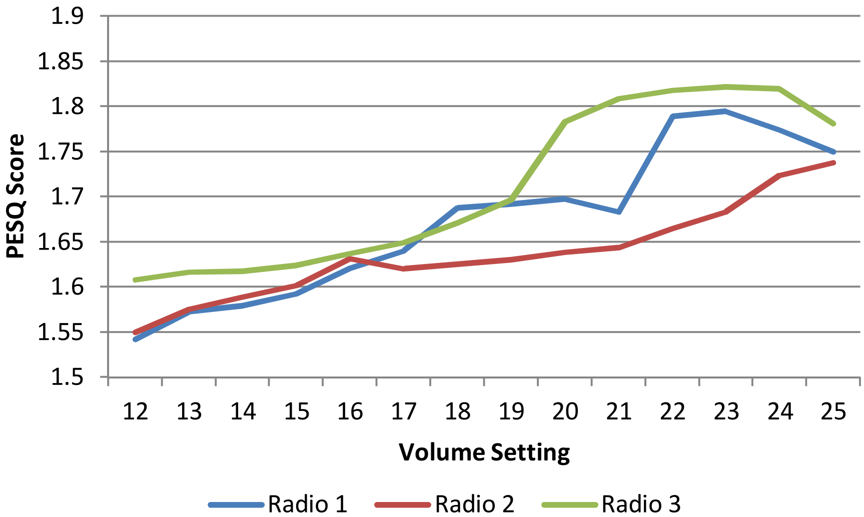


Figure 61 PESQ result for higher volume settings.

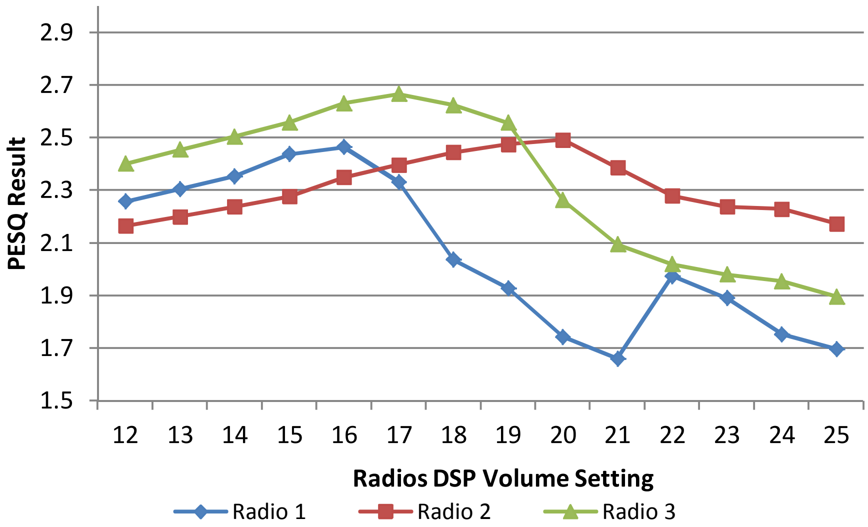


Figure 62: Original PESQ results for the higher volume levels, from the previous section, with no parasitic noise.

Another interesting observation that can be made is the change in position of radio 2 from the radio with the highest PESQ for a clean test signal, to the lowest scoring radio for the signal with parasitic noise. One point that can be made with regards to this is to reiterate the difference in SPL, which was recorded in the section 4.1.2, between the three radios. Because the signal produced by radio two is quieter, the signal to noise ratio of the signal with the noise added will be lower and will therefore culminate in a higher disturbance level for the PESQ algorithm and thus a lower score. Extending this theory it is logical that radio 3 would be the best of the three radios in the high noise environment, as it is also the loudest.

It can also be seen that the peak in PESQ score for radio one at volume level 22 is still present in the background noise tests. Interestingly the sharp reduction in PESQ score seen at volume level 20 for radio three is transformed into a increase in PESQ, suggesting that the mechanism which degrades the audio quality in a quiet environment, improves it in a noisy one.

The next stage of testing involved the measurement of the STI at each of the three radios' volume levels with the addition of the parasitic noise to the recorded test signal. It was noted that once again the intelligibility appears to be quite dependent on the SPL production of the radio. This can be seen by the change in position of radio 2's STI from the test with no noise and the test with the added noise, as shown in Figure 64 and Figure 63 respectively. From the SPL testing in section 4.1.2 it was shown that at volume level 21, radio 1's SPL production peaks and matches radio 3's production before dropping off again. This is reflected in the STI results shown in Figure 63 by the way radio 1 surpasses radio 3 before maximum volume, then drops back below it again.

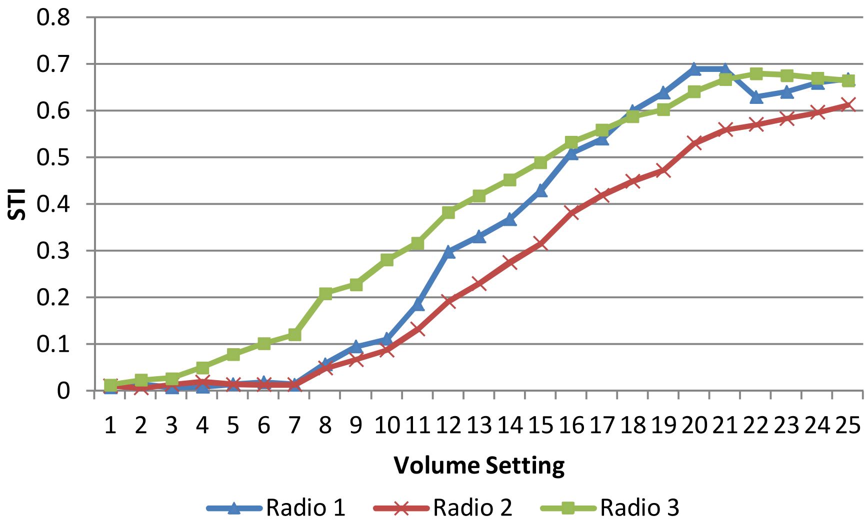


Figure 63: STI of Radio's 1-3 with parasitic noise added.

One of the most important observations to be made from this STI testing is that the STI values for the radios with the noise added are not necessarily always lower than those of the original tests that were conducted in section 4.1.2. This particularly appears to be the case with radios one and three at higher volume levels, with the reverse happening for radio two.

Considering the SPL and PESQ results, it appears that the radios with the higher distortion levels and greater SPL appear to perform better with respect to the STI in the presence of noise.

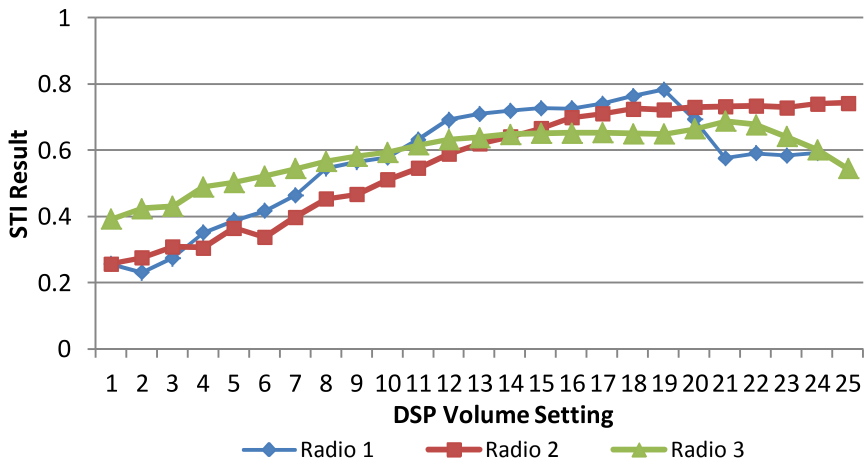


Figure 64: STI as a function of volume level for radio's 1-3.

With the combination of STI and PESQ along with the addition of parasitic noise a useful insight is gained into the effect of SPL and distortion on speech quality and intelligibility in high noise environments.

Radio X testing

The first test to be performed on Radio X was the PESQ score. This was measured at a number of different volume levels with the A weighted SPL being measured at each one. The results showed that radio X produced a similarly shaped PESQ curve to that of radio 3, as shown in Figure 65. Radio X was approximately 1 dB louder than radio 3 and at this volume level had a very slightly higher PESQ score. Below this volume the score is significantly less than that of radio three. The general shapes of the two curves are quite similar.

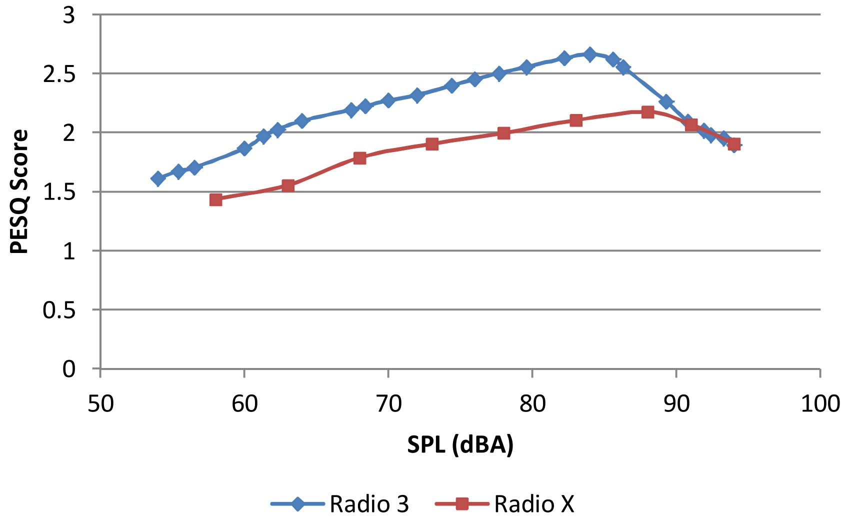


Figure 65: PESQ result of Radio 3 and Radio X with respect to A weighted SPL.

The next test was to compare the STI result of radio X and radio 3, and results of this test are shown in Figure 66. In this test radio X has a significantly higher STI then that of radio three. This supports the theory that intelligibility and quality are not dependant on one another as this radio, when compared with radio three, is better in one area and worse in another.

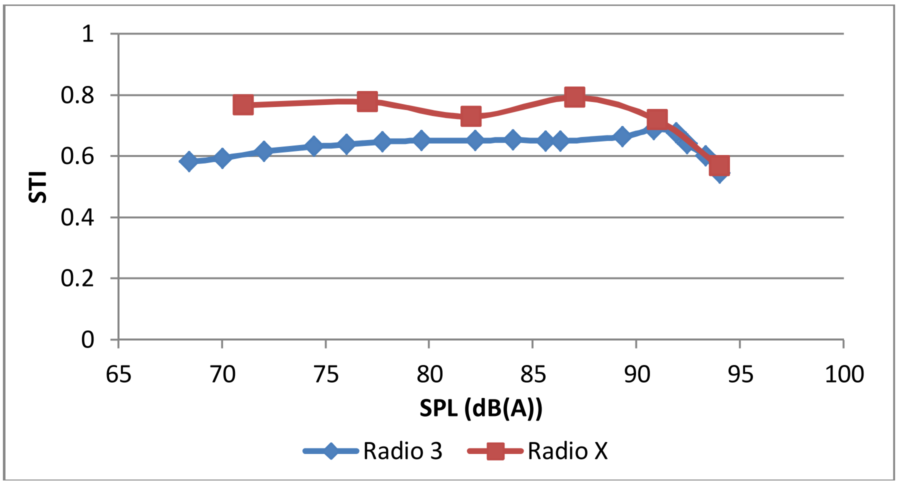


Figure 66: STI result of Radio 3 and Radio X with respect to A weighted SPL.

Finally, to determine how similar radio X is to the other radios, the frequency response was measured using the same method used to produce results in section 4.4. This shows that the radio has a very flat response through the passband region of 300 to 4000 kHz, even flatter then radio 3, which is the flattest of the other three radios. There is also a sharp cut-off point at the end of the pass band, similar to the frequency response of radio 2, as shown in Figure 57. This suggests that there is potentially quite a large amount of distortion present in this signal. This once again gives an example of how distortion can improve intelligibility but be detrimental to quality.

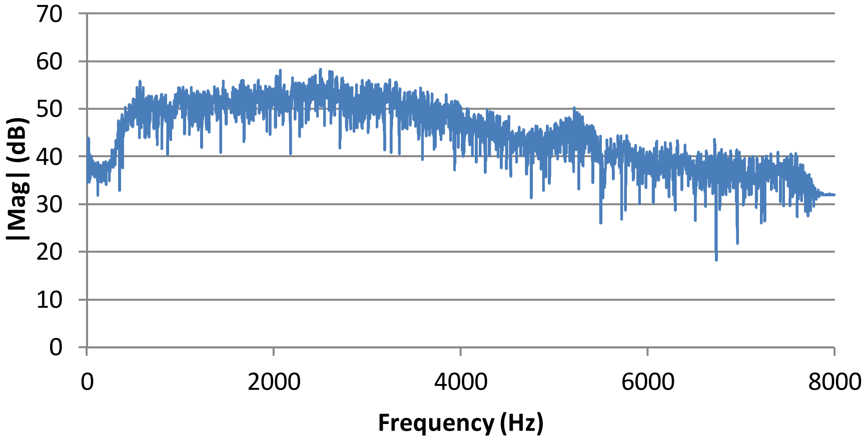


Figure 67: Measured frequency response of radio X at full volume.

Good Quality Large Speaker Test

The final test was to measure the PESQ score and STI for a large, wideband, high power speaker. The results of this test reveal the speaker to have a PESQ score of 3.5 and a STI of 0.83. These results are comparatively high when compared to the radio testing that has been conducted. The reason they are not 4.5 and 1, ie perfect PESQ and STI is that there is no doubt some noise added through the amplification stage of the amplifier, the recording and playback hardware as well as noise induced on the audio cables during the test. Assuming that these parameters remain relatively constant they are not a need for concern as they will reflect evenly on all of the measured results.

6 Audio Quality Optimization

After developing tools for measuring the audio quality and intelligibility of radios, the next logical step was to design a method to improve the audio quality of a system. One such method of achieving this is by altering the frequency response of the system in order to reduce distortion by reducing resonant peaks in the transfer function. This means the frequency response of the radio's audio path would be altered so that when combined with the frequency response of the amplifier and speaker, the net result would be an even spectral distribution.

Correcting the frequency response of a system so that there is an even spectral distribution assumes that an evenly distributed spectrum is the most optimal for high audio quality and intelligibility. From the vowel space and segmental analysis methods described in section 2.3 it was determined that some regions of the frequency spectrum are of more importance than others. Based on this information, the intelligibility of a system could potentially be enhanced by detecting the formant frequencies of the incoming speech signal and amplifying them. This method has already been implemented for public announcement systems in high noise environments such as train stations, [21]. The formant extraction algorithm [20] used for the vowel space analysis described in section 2.3 forms the basis of these formant amplification systems.

This section investigates the feasibility of implementing an automated system that manipulates an equalizer in the radio's audio path. Using the audio quality measures described in sections 2.1 and 2.2, the described system could iteratively change the frequency response of the audio path, using the results of the objective measures as a cost function. Using this method an optimum frequency response could potentially evolve, which would provide optimal audio quality and intelligibility for a particular system.

As part of this investigation a parametric equalizer was implemented in order to investigate the effect of various parameters on the audio quality of the radios. The implementation was designed as a proof of concept rather than a deliverable that could be installed directly onto radios. For this reason only a limited number of band filters were used, in order to reduce potential permutations and thus computation time. The parameters that define the characteristics of the filter were changed iteratively and the resulting PESQ value was measured using a variation of the test setup described in chapter 3.

The hypothesized results were expected to show a varying PESQ score with respect to the parameter being changed and thus a confirmation that this method of optimization shows potential.

Test Setup

In order to conduct this testing a variation of the standard test setup described in chapter 4 was used. This variation involved inserting a Digital Signal Processor (DSP) into the audio line that comes out of the USB sound card and goes to the transmitting radio, as shown in Figure 68. The audio signal was then filtered prior to transmission. Ideally the filtering would occur after the transmission and before the audio signal is amplified, but due to hardware constraints this was not possible.

Due to the number of possible permutations inherently available in this kind of testing only radio 3 was used for testing and only at the maximum volume level. The PESQ algorithm was the only measure of audio quality used for this testing as this testing was only a proof of concept. Other then these variations the test suite was run as per normal with the PESQ test files being played and recorded by the computer and the parametric filter being implemented on the DSP.

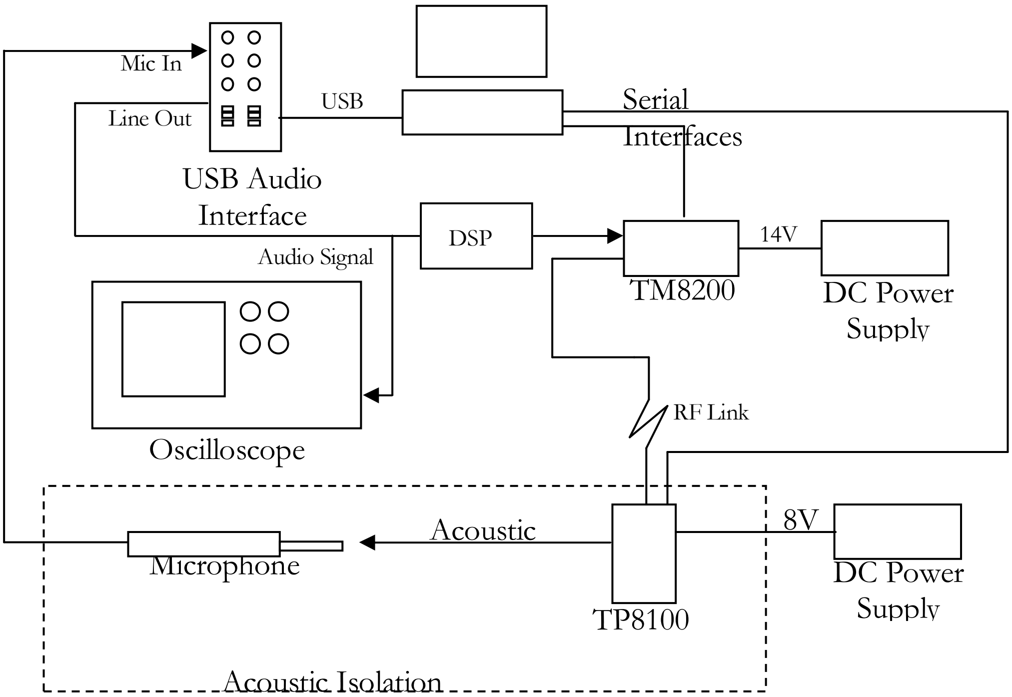


Figure 68: Test setup with DSP implementation of the parametric equaliser inserted into the audio line.

Parametric Equalizer

A parametric filter is a filter that only affects a specific area of the frequency spectrum, having a gain of 0 dB everywhere else. The filter's bandwidth, centre frequency and gain can all be changed by single parameters as the filter is running, making parametric filters extremely versatile and suitable for this type of iterative design.

There are a number of different methods for implementing parametric filters and for this test the design was based on a second order, infinite impulse response, peaking filter [28]. The transfer function of this filter is shown in Equation 16 where ω_0 and $\Delta\omega$ are the normalized centre frequency and bandwidth respectively that describe the filter. A is the cut off level of the band being modified, nominally set to 3 dB.

$$H_{BP}(z) = \frac{\mathrm{C}}{1+\mathrm{C}}\left(\frac{1-z^{-2}}{1-\frac{2\beta}{1+C}z^{-1}+\frac{1-C}{1+C}z^{-2}}\right) \tag{16}$$

where:

$$\beta = \cos(\omega_0), \quad C = (\frac{1}{10^{0.1A}-1})^{0.5}\tan(0.5\Delta\omega)$$

A parametric equalizer is formed by inserting the filter described above into the structure shown in Figure 69, where G is the gain. When G is set to 1 the branch containing the peaking filter structure has no output and the structure has a gain of 0dB across all frequencies. When G is set to a value greater than 1 the structure becomes a boost filter with centre frequency and bandwidth defined by the equations above and a pass band gain set by G. If G is set to a value less than 1 the structure becomes a notch filter.

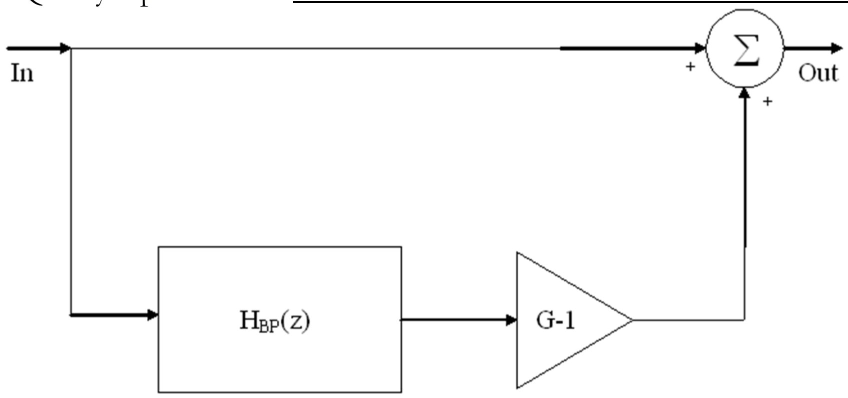


Figure 69; Structure of the second order parametric equaliser.

Test Procedure

The three parameters, centre frequency, bandwidth and gain, were all considered during this testing. In order to determine the effects of each of these, for each test run two were set constant and the third was iterated through a range of values. The centre frequency and bandwidth tests were both run twice, once with the filter in boost mode and once in cut mode. This resulted in 5 test permutations, as described below:

Centre frequency sweep, bandwidth set to 700 Hz

Boost mode, G = 4

Cut mode, G = 0.2

Bandwidth sweep, centre frequency set to 1000 Hz

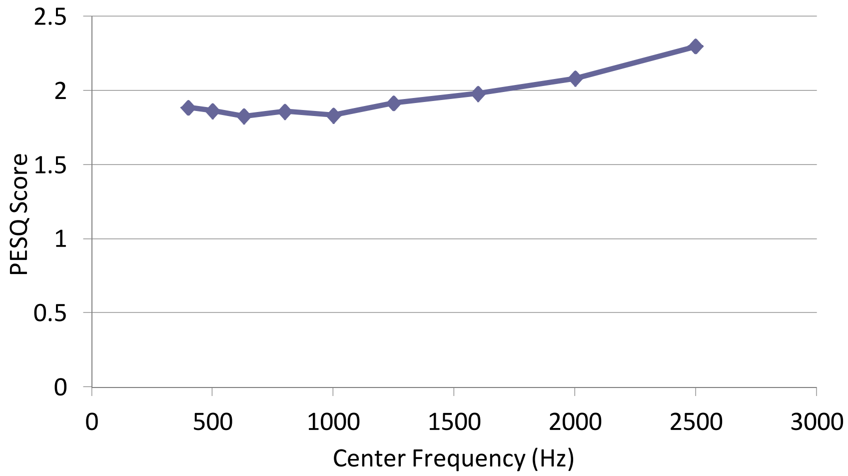
Boost mode G = 4

Cut mode G = 0.2

Gain sweep, centre frequency = 1000 Hz, bandwidth = 700 Hz

Results

The first test to be conducted was the centre frequency tests, where the centre frequency of the parametric amplifier was varied between 400 and 2500 Hz in both cut and boost mode and the PESQ score was measured at each of the settings. The results show that when the filter was located around 1 kHz in boost mode the PESQ result was at its lowest. This result was complemented by the cut mode testing, showing the highest PESQ score when the cut filter was located near 1 kHz. This suggests that the radios are particularly sensitive and highly gained around 1 kHz already and this could potentially be the cause of some of the distortion products.



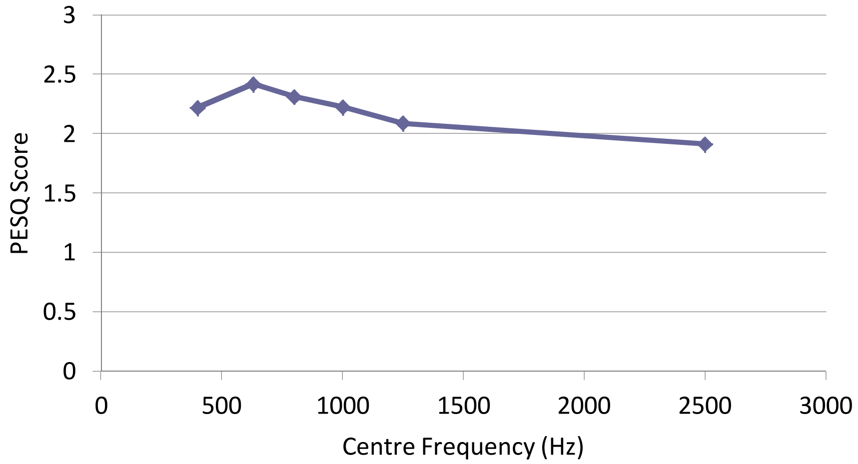


Figure 70 : PESQ as a function of Center Frequency of a parametric equaliser in Boost(Top) and Cut(Bottom) mode.

The results of the bandwidth testing did not reveal particularly useful trends regarding the audio quality and bandwidth of the parametric equalizer. The major shortcoming of these results was that they do not complement one another as the centre frequency testing did. This is a concern as it does not seem logical to have an improvement in PESQ score caused by both cutting and boosting the same frequency. One observation that was noted was that the variation in PESQ score caused by the cut filter was much larger than those caused by the boost filter.

The results from the centre frequency testing, and the fact that the centre frequency of the bandwidth testing is at 1 kHz suggest that an increase in PESQ value would be expected when there is any reduction made near the 1 kHz portion of the spectrum. It is likely that these results were affected by the deviation level of the transmitting radio. Essentially the variations in signal level due to the filtering change the level of the signal going to the transmitting radio, which in turn affects the deviations. These tests potentially reflect the dependency of the PESQ more on the signal strength apparent at the transmitting radio then the bandwidth of the parametric filter.

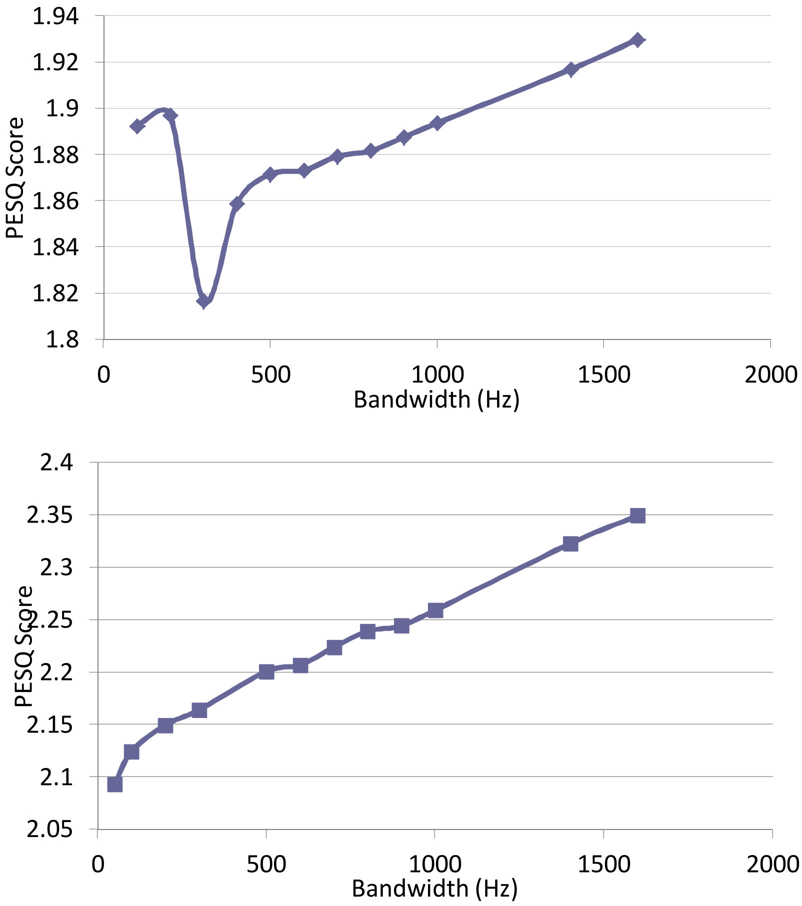


Figure 71: PESQ with respect to bandwidth of a parametric equaliser in Boost (Top) and Cut (Bottom) mode

The final test was to investigate the effects that the gain (G) value of the parametric equalizer structure has on the PESQ score. The centre frequency of the filter was set to 1 kHz and the bandwidth to 700 Hz. The values of G were varied between 0 and 6 with varying step sizes. This test also supports the findings in the first centre frequency testing, which suggest that frequencies around 1 kHz be attenuated. This can seen in Figure 72 specifically at G = 0 where the PESQ score is at its maximum. When G is set to 0 the filter is applying maximum attenuation at the centre frequency, which in this case is 1 kHz. Any value above 0 for G results in a reduction in PESQ score.

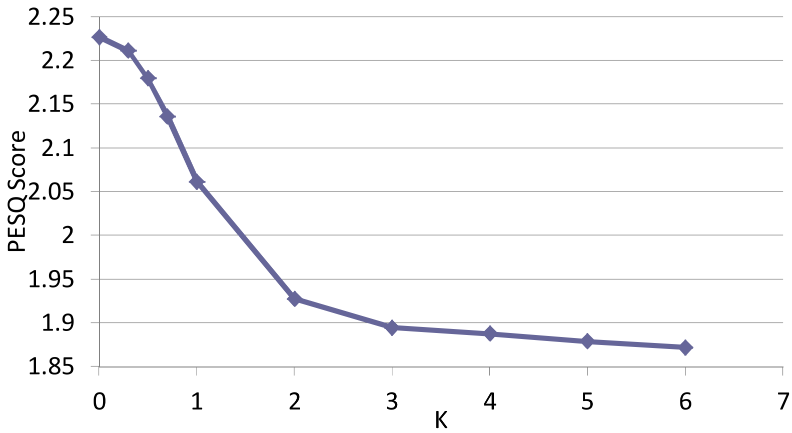


Figure 72: PESQ verse Gain of parametric equaliser.

7 Discussion

Speaker Testing

The effectiveness of a number of speakers as potential replacements for the current speaker was evaluated on the basis of the PESQ and STI analysis methods. The results showed that in the high distortion environment generated by the radio, the cheapest speaker performed surprisingly well, outperforming the other two more expensive speakers at the higher power levels. This was most likely due to the impedance mismatch between the three speakers, causing the higher 16 ohm drivers to get less power from the radio's internal amplifier.

The results from the wide band amplifier testing showed that when sufficient power was supplied to the 16 ohm drivers they could produce substantially higher SPLs than the smaller 8 ohm driver, as well as much better PESQ scores. The importance of matching driver impedance to the radio's amplifier appears to have a greater effect on the audio quality of the system than the power capabilities or quality of the loud speaker being used.

Radio Testing

The results of the PESQ testing of radios 1 through 3 showed that in order to give the results some relevance, the PESQ score needed to be shown with respect to the SPL that the radio could produce at the volume setting being tested. The resulting SPL measurements revealed that radio 3 was the loudest of the radios, and when plotted with respect to SPL, performed the best overall. Radio 2's PESQ scores with relation to discrete volume level were high compared to the other two radios, however, when these scores were normalized for SPL it quickly became apparent that the radio's reduction in SPL at the top volume levels made radio 2 one of the weaker contenders. The high PESQ score at maximum volume is likely related to the reduction in SPL. As the radio is not producing as much power as the other two the distortion products are lower, therefore the radio sounds better though not as loud. The PESQ score for radio 1; when compared with SPL, showed that although it was louder than radio 2 it also had a substantially lower PESQ score.

The STI testing revealed a similar trend, with regards to relating intelligibility to SPL. Radio 2 once again showed promise in the discrete volume level tests; however when it was normalized with SPL and compared to the other radios, radio 3 once again came through as the best, producing the highest SPL with the highest STI. This result however is not as clear cut as the PESQ results. Although radio 3 produces high SPL, its STI is substantially lower than the other two radios until around 90dB.

Relationship between STI and PESQ

Another interesting trend that developed from the comparison of the STI results is the points at which the STI and the PESQ score begin to drop as the radio's volume setting increases. All three radios begin to suffer a reduction in PESQ score at different stages, around 80 dB(A) for radio one and 85 dB(A) for radios two and three. This is not the case for the STI results. The reduction in STI does not occur at all for radio two and for radios one and three it occurs at much higher SPLs, 88 dB(A) for radio one and 92 dB(A) for radio three. When this reduction in STI does begin to occur, the rate of change is much higher than that of the PESQ score. This observation suggests that speech quality and speech intelligibility do not follow the same trends, and a reduction in quality can in fact lead to an improvement in intelligibility.

This observation leads to an interesting area of discussion regarding the roles that SPL and distortion play on the quality and intelligibility of an audio signal. The frequency response measurements that were conducted suggest that radio two has the least amount of distortion at the high volume levels. This would suggest that the PESQ score is quite proportionally dependant on the distortion present, whereas the STI has a more complex relationship: some distortion is beneficial, after which sharp reductions in intelligibility can be expected.

The testing done in conjunction with the validation testing included running the PESQ and STI tests with the addition of a noise signal. The intention of this test was to degrade the radio's output and hence observe the PESQ score and STI reduce as a result. The results obtained from this testing did not correlate with this hypothesis, however. Upon further investigation into the results, it was concluded that the testing methods used were not proven to be invalid. Instead, the results gained from this testing were used to examine the performance of the radios in high noise environments.

The first and most obvious point to make with regards to the results was the change in position of radio two in both the STI and PESQ score with respect to discrete volume setting. The radio went from having the highest PESQ score and STI, to the lowest at the high volume settings in the tests with the added noise. The major difference between radio 2 and the other two radios at these high volume levels was SPL production. This result suggests that one of the most important factors in overcoming noisy environments is volume. This is reiterated in the STI test where the STI of radio 1 surpasses that of radio 3 at volume level 21 then drops back below it again before reaching the maximum volume setting. This reflects the SPL relationship between radio 1 and radio 3 measured in section 4.1.2.

The next interesting phenomenon that the STI and PESQ results both show is the change in score with relation to discrete volume setting. For the first example of this the PESQ result is examined for all three radios. In the original PESQ test, radio three begins suffering a reduction in PESQ score after volume level 20, as does radio two but not as severely. Radio 2 does not suffer a reduction in PESQ for the noisy tests. The PESQ of radio 3 also actually improves at volume level 20 and does not begin to drop until the very last volume setting. Radio one shows a similar shift in PESQ score, with the peak created by the activation of the parametric equalizer still present.

The speech transmission index shows a similar trend. In the original tests radios one and three both suffered from reductions in STI while the volume level was nearing maximum. With the addition of the noise signal the magnitude of the reduction is greatly decreased and at some of the volume settings the STI is actually higher than that of the original test results.

From these observations some conclusions can be drawn with regards to the combination of SPL and distortion and their effects on intelligibility and audio quality, particularly in high noise environments. Firstly, the STI of radio 1 surpasses that of radio 3 in the noisy test, but the PESQ score does not. Radio 1 is never louder then radio three, so the degradation of the signal (that causes radio one to have a lower PESQ score then radio three at volume level 21), actually improves the intelligibility in a noise environment. So much so that it is in fact better than the slightly louder radio 3 as far as intelligibility is concerned.

The second conclusion that can be drawn is with regards to the radio's operation in noisy environments, specifically wide band noise. In the high noise tests the PESQ score is largely dependent on the signal to noise ratio at the recording position, therefore the louder the signal, the better the score. However, the shape of the results generated by the testing with the added noise do not reflect those of the original tests. In fact, when there is a dip in the original tests there is a peak in the noisy tests. This can be explained by masking, the effect where a signal is concealed by another signal. In a normal situation, when the radios volume is increased, the amplitude of the output increases, as do the distortion products, leading to a bigger signal with more disturbances. The addition of these disturbances leads to a lower PESQ score. In a noisy environment the addition of noise masks these distortion products, which simply leaves an increase in signal amplitude and the disturbance level stays the same. Therefore, an increased signal level means higher signal to noise ratio and thus, a higher PESQ score.

In the case of the intelligibility, particularly the situations where the STI value is actually higher in the noisy tests then it is in the quiet test, masking is once again the cause. BronkHorst [29] conducted a review regarding research about the effects of noise on intelligibility. This reveals that noise with a spectrum that correlates with that of the original signal degrades the intelligibility more than noise which has no spectral similarities. For example, it is harder to understand what a speaker is saying when there is another speaker talking beside them than if there was a white noise source. This example can be translated to the radios, where the other speaker is the distortion products generated by the speaker and amplifier, as this has similar spectral content to that of the received speech signal. When the noise is added, the distortion products are masked out by a signal which has less spectral similarity and therefore, the speech is easier to understand. The STI method allows for this phenomenon [29] which makes it a useful measure for determining the intelligibility of radios in noisy environments.

Vowel Space Analysis

The vowel space testing showed that the degradation that occurs to an audio signal during a transmission does affect the formant frequencies of vowel sounds, and hence the distribution of the vowel is a formant 1 formant 2 plane of their vowel space. The results showed that radio 1 had the smallest vowel space across the volume range of the radios with radios 2 and 3 appearing larger than 1 but similar to one another. This supports the findings of the PESQ and STI testing, where radio 2 and 3 were generally better then radio 1 in low noise conditions.

The parameter used to quantify vowel space was the area covered by the triangle that is generated by positioning the three vowels used for testing in a F1 – F2 plane. The variability apparent in these results suggests this is perhaps not the most optimal method. Other methods have been used for quantifying the vowel space. Bradlow[16] uses a method which considers the dispersion for the vowel sounds in the F1-F2 plane from a central point. Turner and Ying-Chiao[14, 18] used 4 vowel sounds to create a four sided shape, the area of which is then used to quantify the vowel space. The papers are looking at the effects of people's vowel space on the intelligibility of their speech. Because in this testing the original input signal is also available another measure could be used which utilizes both input and output signal. For example, the distance between the original signal's vowel distribution and that of the degraded signal could be considered and hence used to quantify the change in the vowel space.

There are a number of possible reasons for the change in vowel space of the vowel sounds after transmission. The most obvious one would be distortion products, as these appear to have the greatest effect on the audio quality and intelligibility. Due to the high level of distortion at some of the higher volume levels, the distortion products may have higher amplitudes then the original signal, which means the distortion product, could then become the formant frequency.

Some of the formant results however showed a high level of consistency across the entire volume range with no noticeable change at the high volume levels where the distortion products begin to become apparent. This suggests that the distortion is not the only cause of the change in vowel space. Another factor is potentially the bandwidth of the transmission system, being 300 Hz to 3800 Hz. Because some of the vowel sounds have been shown to have formant frequencies below or close to 300 Hz these spectral peaks are likely to be filtered out by the sampling process performed during a normal radio transmission. This will cause at the very least a reduction in the amplitude of the spectral peak with a possibility of complete deletion. In this case another spectral peak will need to take its place, hence leading to a change in vowel space.

Segmental Analysis

The majority of the results that were calculated as a part of the segmental analysis showed a large amount of variation with respect to volume level and little to no correlation with the results obtained from the PESQ and STI testing. The effect of the radio transmission on the parameters under investigation was quite drastic, with significant variation generally occurring across the majority of the volume range.

Both Singer Power Ratio and F1 F2 slope are measures of spectral distribution and because the frequency response of the radios is not flat it is expected that there would be some changes. The interesting point to note is that there does not seem to be a change in any of the measures with respect to volume, particularly to highlight the addition of distortion products. This either suggests that the parameters being measured are not affected by distortion, or the results are erroneous. Due to the fact that the frequency response of the radios changes with respect to volume level, and that the results of the segmental testing do not show any dependency on this variation, as well as the sporadic nature of a large percentage of the test results, it was determined that the accuracy of the segmental analysis was questionable.

Optimization equalizer

The optimization equalizer had varying levels of success across the tests that were run. The results, which showed some correlation with one other, suggested that the frequency response should be adjusted to reduce the gain in the frequency spectrum at around 1 kHz. This was verified by centre frequency testing in both cut and boost mode, as well as the gain testing, where the highest PESQ score was obtained when the equalizer was set to have maximum attenuation at 1 kHz.

Ideally this test would be performed at the output stage of the receiving radio's amplifier, essentially tuning the frequency response of the audio path to that of the enclosures and speaker resonant responses. The testing that was completed was done in order to prove the concept of an automated equalizer. In the future the filter could be implemented into a genetic search algorithm in order to cover a wider selection of the permutations of the filter's parameters, while also incorporating finer frequency resolution.

Analysis Methods Review

The analysis methods used for the high level subjective testing were PESQ and STI. These measures were used to determine the speech quality and speech intelligibility respectively. Both of these measures showed results, that correlate with existing subjective opinions about the radios. The advantage of these high level methods is they can be used to assess the changes caused by any number of parameters controlling the radios, as was done in the optimization equalizer section.

Tests measuring the STI use a non-speech test signal. This is appropriate for analogue based radios as there is no speech specific processing happening. This is not the case however for digital radios, which incorporate the use of vocoders to encode the speech into a digital form. The encoding method uses characteristics of the human speech signal to optimize the transformation. The STI test signals do not have these characteristics so they will be poorly represented by the vocoder process. The STI method can be used for testing the analogue subsystems within these radios as an overall test, hence it is not transferable to digital mobile radios.

The PESQ system does not have this shortcoming as the test signals are comprised of recorded speech, and can therefore be transmitted by digital radios. The system is also only validated for vocoders working at rates higher than 4 kbits, which is higher than some of the rates that are recommended for more recent digital standards. Another shortcoming of the PESQ system is the test signal length, which needed to be greater than 8 seconds. This makes it difficult to measure the effects from perceptual based signal processing methods. This is because the processing will only affect a small section of the speech signal, and the magnitude of the effects will be scaled down by the averaging of the PESQ score with the rest of the signal. Overall PESQ is a very useful and robust analysis tool that can be used in a variety of configurations in order to determine the effect a number of parameters and different environments have on speech quality.

The vowel space and segmental based analysis used in this testing pave the way for new possibilities with regards to signal processing algorithms and speech reproduction. By examining the way speech is produced and hence understood, more efficient methods of improving audio quality and intelligibility can be implemented. An example of one of these methods is the public announcement system in a train station [21], where the formant frequencies are found during the amplification of a speech signal, and hence boosted. This has been shown to produce an improvement in intelligibility, and may be more attractive than increasing the loudness of a system. A similar based system could potentially be implemented on the radios, biasing the importance of the formant frequencies during transmission, resulting in an increase in intelligibility, without an increase in required SPL production.

New standards for measuring audio quality in these new digital environments already exist. ETSI ES 202 738 [30] describes test setups and analysis methods for determining the transmission qualities of Voice over Internet Protocol systems.

8 Conclusion

The Test Suite

As part of this project a functional test suite was developed that can be used for measuring the audio quality of various radios. This system utilizes the PESQ and STI method for measuring the speech quality and speech intelligibility respectively. These methods of analysis have shown results that generally agree with subjective results for the subject radios.

Speech Quality versus Speech Intelligibility

The test suite developed was used to investigate the audio quality of a number of different radios. Three radios were from the same series but from different stages of development. This testing revealed that the latest and most powerful of the radios was the best performing in terms of both quality and intelligibility. Speech quality and intelligibility were both shown to not necessarily always complement each other. The testing suggested that speech quality is more closely related to distortion than intelligibility. Hence, an increase in distortion can lead to a reduction in quality while also providing an increase in intelligibility. As the distortion level increases however, the intelligibility peaks and then degrades quickly.

Validation tests initially showed the importance of SPL in noisy situations but also the improvements in intelligibility and quality that can be made with the addition of distortion. The effect that caused this phenomenon is referred to as “masking”, and occurs when the background noise signal has no spectral similarities to that of the speech signal, and conceals the distortion products generated by the radio. Using the knowledge gained about distortion's effect on intelligibility and quality, particularly in noisy environments, new methods for configuring the radio's volume and distortion properties could be applied.

Vowel Space

The testing that was performed in order to determine the effect of radio transmission on the vowel space of a vowel sound showed results of variable relevance. The method used to quantify vowel space is potentially at fault for the large amount of variation that was seen with respect to volume level for the calculated vowel space results. A new method could be implemented that considered the change in vowel distribution between the original and degraded signal.

The formant extraction algorithm performed reasonably well, greatly reducing the analysis time required for the vowel space testing and hence making it more feasible. There were a number of recordings for which the formant extraction system failed to find the correct formant values, but these were usually extremely distorted signals with ill defined spectral peaks.

This form of testing shows potential in the development of new signal processing methods that could help improve the audio quality of transmission systems. A formant tracking algorithm could be implemented within the DSP on the radio and used to boost the formant sounds prior to transmission, improving intelligibility.

Segmental Analysis

The segmental analysis showed results with significant variations across the volume range tested. The variability and independencies of the results with respect to changes in the audio path of the radio suggest that the segmental analysis methods used were not valid indicators of speech quality. Also the radios which offered the results closest to those of the original test signal did not show the best results from the PESQ and STI testing, suggesting that this method of analysis is not effective for quantifying the audio quality of radios.

Automated Equalizer

The final part of the project was to determine the feasibility of an equalizer that optimizes the frequency response of the radio based on iterative testing involving the test suite. The testing performed under this section involved iteratively changing the parameters of a parametric equalizer and determining the effect these changes had on the audio quality. The results that were gathered as part of this testing showed that this type of development is feasible and potentially quite rewarding in terms of improvement in audio quality. Even though the test setup was limited by existing functionality and only a small number of test iterations were run, the results showed some trends that already suggest some changes that could be made to improve the audio quality of the radios.

Overall the test suite that was developed proved to be a very versatile and informative tool. The results that have been obtained from this project have given some insight into the future development of audio processing systems in the radios, and have opened the doors for new possibilities in terms of new optimization algorithms and processing.

A key step in the development of the new optimization systems will be the combining of the PESQ and STI measures with the lower level methods, particularly vowel space. By looking at the effect the radios have on speech signals at a perceptual level, more efficient methods of improving speech quality and intelligibility can be implemented.

9 References

[1] ITU. P.862, "**Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,**" in *Methods for objective and subjective assessment of quality*, 2001.

[2] IEC. 60268-16, "**Objective rating of speech intelligibility by speech transmission index,**" 2003.

[3] ANSI/ASA S3.2-2009, "**Method for Measuring the Intelligibility of Speech over Communication Systems,**" 2009.

[4] ITU. P.862.3, "**Application guide for objective quality measurement based on Recommendations P.862, P.862.1 and P.862.2,**" in *Methods for objective and subjective assessment of quality*, 2008.

[5] ITU. P.800.1, "**Mean Opinion Score (MOS) terminology,**" in *Methods for objective and subjective assessment of quality*, 2006.

[6] ITU. P.800, "**Method for Subjective Determination of Transmission Quality,**" in *Methods for objective and subjective assessment of quality*, 1996.

[7] A. E. Mahdi, "**Voice quality measurement in modern telecommunication networks,**" Piscataway, NJ, USA, 2007, pp. 25-32.

[8] ITU. P861, "**Objective quality measurement of telephone-band (300-3400 Hz) speech codecs,**" 1998.

[9] E. Zwicker, "**Subdivision of the Audible Frequency Range into Critical Bands (Frequenzgruppen),**" *The Journal of the Acoustical Society of America,* vol. 33, p. 248, 1961.

[10] S. S. Stevens, "**The Measurement of Loudness,**" *The Journal of the Acoustical Society of America,* vol. 27, pp. 815-829, 1955.

[11] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "**Perceptual evaluation of speech quality (PESQ) - A new method for speech quality assessment of telephone networks and codecs,**" Salt Lake, UT, United states, 2001, pp. 749-752.

[12] H. J. M. Steeneken and T. Houtgast, "**Basics of the STI-measuring method,**" http://www.steeneken.nl/sti/Basics_STI_web-site.pdf.

[13] H. J. M. Steeneken and T. Houtgast, "**A physical method for measuring speech-transmission quality,**" *The Journal of the Acoustical Society of America,* vol. 67, pp. 318-326, 1980.

[14] T. Ying-Chiao, W. Gary, and I. Kamran, "**The effect of intertalker speech rate variation on acoustic vowel space,**" *The Journal of the Acoustical Society of America,* vol. 119, pp. 1074-1082, 2006.

[15] L. Huei-Mei, T. Feng-Ming, and K. K. Patricia, "**The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy,**" *The Journal of the Acoustical Society of America,* vol. 117, pp. 3879-3889, 2005.

[16] A. R. Bradlow, G. M. Torretta, and D. B. Pisoni, "**Intelligibility of normal speech I: Global and fine-grained acoutic-phonetic taker characteristics,**" *Speech Communication* vol. 20, pp. 255-272 1996.

[17] J. G. Proakis and D. G. Manolakis, ***Digital Signal Processing***, 4th ed.: Prentice-Hall, 2007.

[18] G. S. Turner, K. Tjaden, and G. Weismer, "**The Influence of Speaking Rate on Vowel Space and Speech Intelligibility for Individuals With Amyotrophic Lateral Sclerosis,**" *J Speech Hear Res,* vol. 38, pp. 1001-1013, October 1, 1995 1995.

[19] R. J. Baken, ***Clinical Measurement of Speech and Voice.***: PRO-ED, Inc, 1987.

[20] S. McCandless, "**An algorithm for automatic formant extraction using linear prediction spectra,**" *IEEE Transactions on Acoustics, Speech and Signal Processing,* vol. 22, pp. 135-141, 1974.

[21] *H. Brouckxon*, *W. Verhelst*, and *B. D. Schuymer*, "**Time and Frequency Dependent Amplification for Speech Intelligibility Enhancement in Noisy Environments,**" in *Interspeech 2008*, 2008, pp. 557-560.

[22] R. E. Stone, J. T. F. Cleveland, P. J. Sunberg, and J. Prokop, "**Areodynamic and Acoustical Measures of Speech, Operatic, and Broadway Vocal Styles in a Professional Female Singer,**" *Journal of Voice,* vol. 17, pp. 283 - 297, 2003.

[23] W. Whyte, "**Computer Controlled Data Interface Protocol Specification** " CCDI version 3.06 Tait Electronics Ltd, 2007.

[24] Lexicon, "**Lexicon Lambda Data Sheet,**" www.lexicon.com, 2008.

[25] Behringer, "**ECM8000 Tehnical Specifications,**" http://www.behringerdownload.de/ECM8000/ECM8000_C_Specs.pdf, 2000.

[26] Mechanical Engineering Department, "**Laboratories,**" University of Canterbury, http://www.mech.canterbury.ac.nz/, 2009.

[27] C. Hansen, ***Noise control: from concept to application*** Spon Text, 2005.

[28] A. Ambardar, ***Analog and Digital Signal Processing*** Second ed.: Brooks/Cole Publishing 1999.

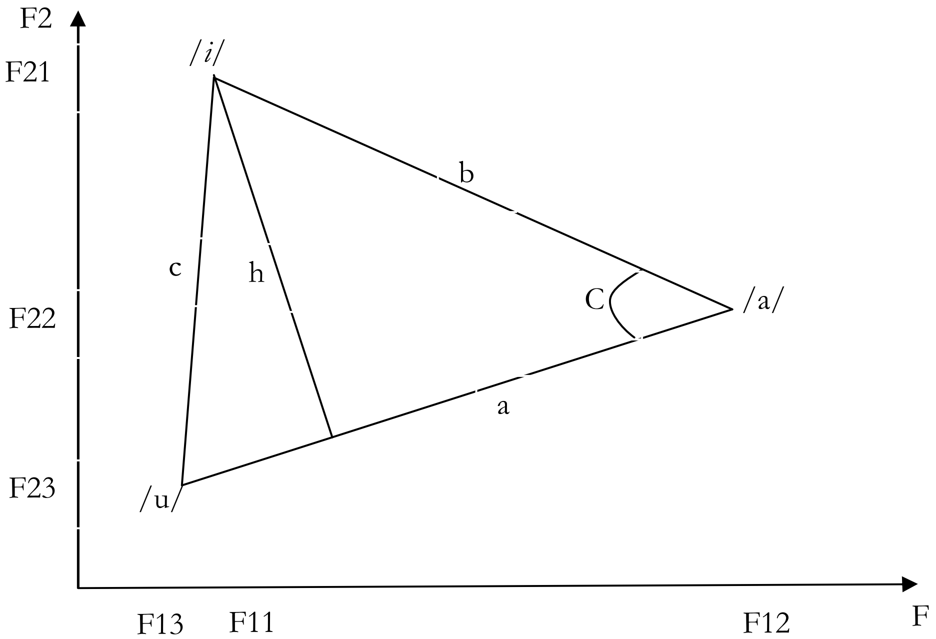
[29] A. W.Bronkhorst, "**The Cocktail Party Phenomenon: A Review of Reseach in Speech Intelligibility in Multi-Talker Condtions,**" *Acustica* vol. 86, pp. 117-128, 2000.

[30] ETSI. ES 202 740, "**Speech and multimedia Transmission Quality (STQ); Transmission requirements for wideband VoIP loudspeaking and handsfree terminals from a QoS perspective as perceived by the user,**" 2009.

10 Appendix

Appendix A

Area of a Triangle.



The length of h can be determined by the sine identity using sides a and b along with angle C

$$h = b\ sin(C)$$

Then using the equation to find the area of a triangle

$$Area = \frac{ah}{2}$$

Substituting in the two equations gives

$$Area = \frac{a\ b\ sin(C)}{2}$$

To find the angle C the cosine rule is used;

$$Cos(C) = \frac{a^2 + \ b^2 - \ c^2}{2ab}$$

Substituting these last two equations reveals

$$Area = \frac{ab}{2} sin(\ cos^{-1}\left(\frac{a^2 + \ b^2 - \ c^2}{2ab}\right))$$

Finally the definition of a,b and c is as follows

$$a = \sqrt{(F12 - F13)^2 + \ (F22 - F23)^2}$$

$$b = \sqrt{(F12 - F11)^2 + \ (F21 - F22)^2}$$

$$c = \sqrt{(F11 - F13)^2 + \ (F21 - F23)^2}$$

Appendix B

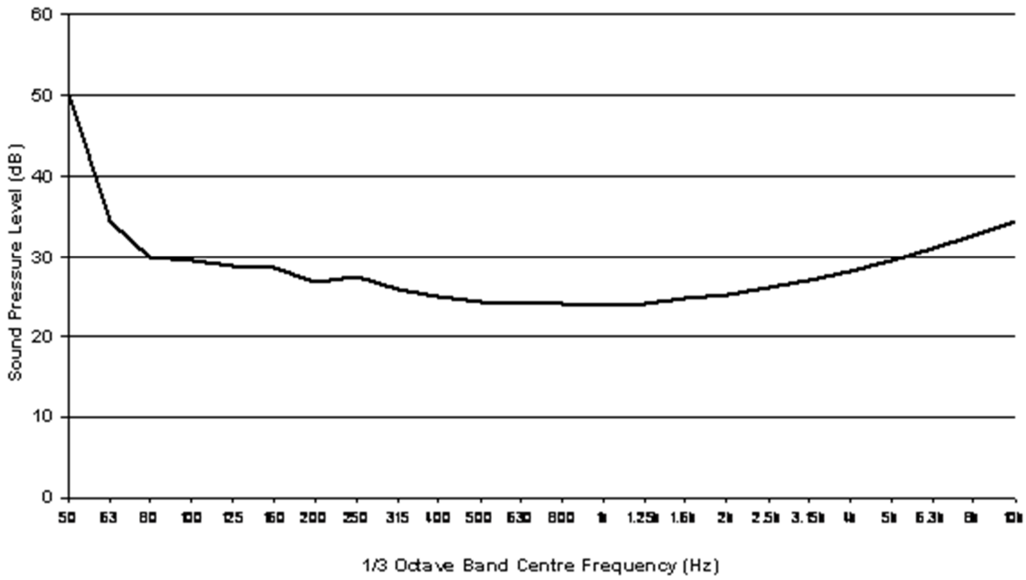
LPC Algorithm

```
def AutoCorrelation(RealSig1, Order):
    RealSig = RealSig1/100
    R = []
    Ns = float(len(RealSig))
    lag = 0
    while lag <= (Order):
        R.append(numpy.correlate(RealSig[0:Ns-lag],
                    RealSig[lag:Ns])[0])
        lag += 1
    R = numpy.array(R)
    R = R / Ns
    R = R*10000
    return R

def LevinsonDurbin(RealSig, Order):
    a_m = numpy.zeros((1, Order + 1))[0]
    a_mm1 = numpy.zeros((1, Order +  1))[0]
    R = AutoCorrelation(RealSig, Order)
    a_m[0] = 1.0
    a_m[1] = -1.0 * R[1] / R[0]
    m = 2
    while m <= Order:
        a_mm1 = numpy.array(a_m.tolist())
        Cor = a_mm1[1:m].tolist()
        Cor.reverse()
        Cor = numpy.array(Cor)
        a_m[m] = -1 * ( (R[m] + numpy.correlate(R[1:m],
                            Cor)[0])/
                (R[0] + numpy.correlate(R[1:m],
                        a_mm1[1:m])[0]))
        k = 1
        while k < m:
            a_m[k] = a_mm1[k] + a_m[m]*a_mm1[m - k]
            k += 1
        m += 1
    return a_m
```

Appendix C

Background Noise Levels in Anechoic Room



T_{20} Reverberation Times of Anechoic Room